

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 20:55:17 ; Search time 2937 Seconds

(without alignments)  
15879.120 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1140

Sequence: 1 atgaccagcgtcaagtgcctga.....taacatcggtcgcgaagtga 1140

Scoring table:

OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_ncg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_pn:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1140	100.0	1140	6	AX548642
2	1140	100.0	6334	6	AY178757 Rhodococc
3	1140	100.0	6334	6	AX548646
4	1140	100.0	9652	6	AX548648
5	1140	100.0	11241	6	AX548647
6	1140	100.0	11241	12	AY180162
7	22	1.9	202897	2	AC133415
8	22	1.9	224046	2	AC132319
9	22	1.9	228864	2	AC103497
10	21	1.8	147405	2	AF005161
11	21	1.8	155547	2	AF005799
12	20	1.8	20	6	AX548660
13	20	1.8	167	6	AX049359
14	20	1.8	1402	9	AY227111
15	20	1.8	1420	9	AY227112
16	20	1.8	1428	6	AX049362
17	20	1.8	1501	6	AX717763
18	20	1.8	1513	6	AX717761
19	20	1.8	1961	6	AX714753
20	20	1.8	1961	6	AK057340
21	20	1.8	3114	6	AX686980
22	20	1.8	5585	1	AF516142
23	20	1.8	10888	1	AE009218
24	20	1.8	11693	1	AE008185
25	20	1.8	89319	9	HSBRK29F11
26	20	1.8	94625	9	AC011337
27	20	1.8	121727	2	AC119124
28	20	1.8	133614	2	AC025433
29	20	1.8	154520	10	AC132098
30	20	1.8	159562	9	AC034205
31	20	1.8	181772	2	AC007689
32	20	1.8	201659	2	AC120558
33	20	1.8	208523	2	AC113512
34	20	1.8	241150	2	AC105884
35	20	1.8	305520	1	AE016780
36	19	1.7	882	8	SAU25696
37	19	1.7	891	4	OCABP1
38	19	1.7	1092	9	HSHBRG1
39	19	1.7	1192	9	HSHBRG
40	19	1.7	1246	4	OCU57553
41	19	1.7	3390	4	AF144711
42	19	1.7	3810	6	AX411065
43	19	1.7	3810	9	HSHBRG
44	19	1.7	4579	14	SCSCCATRNA
45	19	1.7	4579	14	SCSCCATRNA

## ALIGNMENTS

RESULT 1  
LOCUS AX548642  
DEFINITION Sequence 1 from Patent WO02055709.  
ACCESSION AX548642  
VERSION AX548642.1 GI:25813612  
KEYWORDS  
SOURCE  
ORGANISM  
Rhodococcus erythropolis  
Rhodococcus erythropolis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Nocardiaceae; Rhodococcus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
1  
Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.  
Rhodococcus cloning and expression vectors  
Patent: NO 02055709-A 1 18-JUL-2002

FEATURES  
 source  
 E.I. DU PONT DE NEMOURS AND COMPANY (US)  
 Location/Qualifiers  
 1. 1140  
 /organism="Rhodococcus erythropolis"  
 /mol\_type="genomic DNA"  
 /strain="AN12"  
 /db\_xref="taxon:1833"

BASE COUNT 241 a 287 c 347 g 265 t

ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 1140;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ATGACCGAGCTAGTGTGTAACACCTTCCGCGCAAGACCGGCTCCCGTCTCGTGTG 60
Db 1 ATGACCGAGCTAGTGTGTAACACCTTCCGCGCAAGACCGGCTCCCGTCTCGTGTG 60
Qy 61 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCACCAGTCA 120
Db 61 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCACCAGTCA 120
Qy 121 GAAACATTTAAACGCTGTGCGCGCGCAATTTCTGGCGTGAACGCGTGAACATTGTCAAC 180
Db 121 GAAACATTTAAACGCTGTGCGCGCGCAATTTCTGGCGTGAACGCGTGAACATTGTCAAC 180
Qy 181 GGTCCGAAAGGTTCTGGAATTCGAGGCGCTTGCTGCGGAAAGGCGTGAATTCGCCC 240
Db 181 GGTCCGAAAGGTTCTGGAATTCGAGGCGCTTGCTGCGGAAAGGCGTGAATTCGCCC 240
Qy 241 TGCCTGCGGAAAGTCTGCTGACATCTGACAGACCAAAATTTCTCAAGTTGTTGCTCAT 300
Db 241 TGCCTGCGGAAAGTCTGCTGACATCTGACAGACCAAAATTTCTCAAGTTGTTGCTCAT 300
Qy 301 CAATCGCGAATGATCTGTTGCGATGATGACATGACATGCGCATACAGCTGTGAC 360
Db 301 CAATCGCGAATGATCTGTTGCGATGATGACATGACATGCGCATACAGCTGTGAC 360
Qy 361 CGGCTCCAGACCTTATGACCTTTCGCGACCTTGGCAAGCTGTGACCAACGGTCT 420
Db 361 CGGCTCCAGACCTTATGACCTTTCGCGACCTTGGCAAGCTGTGACCAACGGTCT 420
Qy 421 CGTTGCGGTACGGAACGTGAATTAAGCGCTGCGACGATTAAGCTGTGACATC 480
Db 421 CGTTGCGGTACGGAACGTGAATTAAGCGCTGCGACGATTAAGCTGTGACATC 480
Qy 481 ACTACCGAAAAAAGCGCTGCGACGATCAAGCTTCAAGCGCTACTCATGTTCAAGTGTGAC 540
Db 481 ACTACCGAAAAAAGCGCTGCGACGATCAAGCTTCAAGCGCTACTCATGTTCAAGTGTGAC 540
Qy 541 GTGAGTGAACATCTCTGAACTCTTCTCGATGCGATGTTCCATCGGTGACCTTCCAAA 600
Db 541 GTGAGTGAACATCTCTGAACTCTTCTCGATGCGATGTTCCATCGGTGACCTTCCAAA 600
Qy 601 CTGCTATCTCTGGGATTTGTCGCGCACTACGTAATTCGAGTGTGCTCGATGACGAAG 660
Db 601 CTGCTATCTCTGGGATTTGTCGCGCACTACGTAATTCGAGTGTGCTCGATGACGAAG 660
Qy 661 ATCGCGGTGAAGCTGATCAAGTTCTGCTGCGATCTGACGAAGATTCGATCTGGCGTT 720
Db 661 ATCGCGGTGAAGCTGATCAAGTTCTGCTGCGATCTGACGAAGATTCGATCTGGCGTT 720
Qy 721 GGTATGAGAGTTGATGATGCGGCAAGGAAATGATGCAATGCGACCTCTCG 780
Db 721 GGTATGAGAGTTGATGATGCGGCAAGGAAATGATGCAATGCGACCTCTCG 780
Qy 781 GAAATCGCTGTTGATGATGCGGCGGGAATCCAAAGGTTGGAATCTGTGCGGAAATTT 840
Db 781 GAAATCGCTGTTGATGATGCGGCGGGAATCCAAAGGTTGGAATCTGTGCGGAAATTT 840
Qy 841 GAGTTGTTGATGATGCGGCGGCAATGCGGTGTTCCGTGATTCGTCGCGGAGCT 900
Db 841 GAGTTGTTGATGATGCGGCGGCAATGCGGTGTTCCGTGATTCGTCGCGGAGCT 900

```

Qy 901 GGTCTTGGGCGAGAACTTAACAGATGCTAGATGTTGAGCAGAGAAATCTCCCGGTC 960  
 Db 901 GGTCTTGGGCGAGAACTTAACAGATGCTAGATGTTGAGCAGAGAAATCTCCCGGTC 960  
 Qy 961 ATGTTGCGATCATTTCCGCGCGATCGTGAATGATGATTCGACCTTGGCGCTTAAGTC 1020  
 Db 961 ATGTTGCGATCATTTCCGCGCGATCGTGAATGATGATTCGACCTTGGCGCTTAAGTC 1020  
 Qy 1021 TTGCGGAGATCTCGGATCTCGTCAAGCTGCGCGAGCTTGGGAAATCTTGTGATCAC 1080  
 Db 1021 TTGCGGAGATCTCGGATCTCGTCAAGCTGCGCGAGCTTGGGAAATCTTGTGATCAC 1080  
 Qy 1081 TTGCATTATCGATTCGCGCGAGGATGTCGCGAGCCCGCATTAATTCGATTCGCAAGTGA 1140  
 Db 1081 TTGCATTATCGATTCGCGCGAGGATGTCGCGAGCCCGCATTAATTCGATTCGCAAGTGA 1140

RESULT 2  
 AY178757/c 6334 bp DNA linear BCF 01-MAR-2003  
 LOCUS  
 DEFINITION Rhodococcus erythropolis cryptic plasmid pAN12, complete sequence.  
 ACCESSION AY178757  
 VERSION AY178757.1 GI:28628256  
 KEYWORDS  
 SOURCE Rhodococcus erythropolis  
 ORGANISM Rhodococcus erythropolis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Nocardiaceae; Rhodococcus.  
 1 (bases 1 to 6334)  
 Kostichka, K., Tao, L., Brammucc, M., Tomb, J.-F., Nagarajan, V. and  
 Cheng, Q.  
 A small cryptic plasmid from Rhodococcus erythropolis:  
 Characterization and utility for gene expression  
 Unpublished  
 2 (bases 1 to 6334)  
 Kostichka, K., Tao, L., Brammucc, M., Tomb, J.-F., Nagarajan, V. and  
 Cheng, Q.  
 Direct Submission  
 Submitted (11-NOV-2002) CRD, E.I. Dupont de Nemours Inc.,  
 Experimental Station, Wilmington, DE 19880-0328, USA

FEATURES  
 source  
 1. 6334  
 /organism="Rhodococcus erythropolis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1833"  
 /plasmid="pAN12"  
 /note="cryptic plasmid"  
 /complement (1912..3051)  
 /gene="rep"  
 /complement (1912..3051)  
 /gene="rep"  
 /note="Rep"  
 /codon\_start=1  
 /transl\_table=11  
 /product="repllication protein"  
 /protein\_id="AA031602.1"  
 /db\_xref="GI:28628257"

gene  
 translation="MTSVSAKHLSGDRPEVLYSSDKGTRHRLRPEKQITTSRTN  
 ACGRPIISGVNAGTIVNGPKSGFGGLRSCKGWI CPCAGKVAHRADEISQVVAHQL  
 GTGVSVMVTHMTHTAGORLHDHTGLISAAMKATNGRRTREMYGCDGYRAVAVI  
 THKNGKMHVTHALMLFSGVSENITLESFSDAMPDRTSKLVISGFAPAPRNSGGLDV  
 RKIAGEADQVLAAYLTRIASGVMAVYSGSGKSGRHNRAPEIAYDAVGDQALEL  
 WRPEFGSKMRAIAMSRLGRARAGAGAEITDQIYVEGESSAVVMVALIIPRSMWIR  
 TCAPIYFGEITLGVENGATVEMRDHLHTPLPADVPRPITISVRK"

CDS  
 gene  
 complement (4288..5178)  
 /gene="div"  
 /complement (4288..5178)  
 /note="div"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative cell division protein"



```
/protein_id="AA031603.1"
/db_xref="GI:28628258"
/translation="MDQTDPIPIAIGMNELAOPVWVDIAKDAHMLIOGKSGKSQC
TYMLAQAGSNPAVRVGVDPISVLPAFVHRPAEPNLEIANDPKYLRIQFVKA
ESDRRIECFMDRIIDKISLESPALPLILAEFPDILIEGAOFPDNLKPADRVAP
RTSLIVROIAAOSAKAGIYMLILAGRAEISVGNASRNPAYVMTIRPDESVKHL
PNATPEECALVEGFVPGQGFDPGRKQIRKIRVGEISTIASIYENADLATEALIN
IDRAQRMITASEPFLHLDIG"
BATCH COUNT 1386 a 1807 c 1681 g 1460 t
ORIGIN

Query Match 100.0%; Score 1140; DB 1; Length 6334;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGTGAACAACCTTCCGCAAAAGACCGGCTCCCGTCTGATGTCG 60
Db 3051 ATGACGAGCGTAAGTGTGTGAACAACCTTCCGCAAAAGACCGGCTCCCGTCTGATGTCG 2992

QY 61 TCCGATTAAGCGCGCATCCGCGCAAGACCTGCGAACCCAACTTCAACAAATCACACGTC 120
Db 2991 TCCGATTAAGCGCGCATCCGCGCAAGACCTGCGAACCCAACTTCAACAAATCACACGTC 2932

QY 121 GAAACATTTAAAGCCCTGTGCGCGCGCATTTCTGGCGGTGAACGCTGTGACCATTTGTCAAC 180
Db 2931 GAAACATTTAAAGCCCTGTGCGCGCGCATTTCTGGCGGTGAACGCTGTGACCATTTGTCAAC 2872

QY 181 GGTCCGAAAGGTTCTGGATTCTGAGAGGCTTCTGTTCTCGCGAAAGGCTGGATCTGCCCC 240
Db 2871 GGTCCGAAAGGTTCTGGATTCTGAGAGGCTTCTGTTCTCGCGAAAGGCTGGATCTGCCCC 2812

QY 241 TGTGTGCGGAAAAGTGTGTGACATCTGTGACAGACGAAATTTCTCAAGTTGTGTCAT 300
Db 2811 TGTGTGCGGAAAAGTGTGTGACATCTGTGACAGACGAAATTTCTCAAGTTGTGTCAT 2752

QY 301 CAACCTGGGACTGGATCTGTTGGATGTGTGACATGACATGCGCCATACAGCTGGTCAG 360
Db 2751 CAACCTGGGACTGGATCTGTTGGATGTGTGACATGACATGCGCCATACAGCTGGTCAG 2692

QY 361 CGGCTCCAGACCTATGACATGACATTTGGGACGCTGGAAGCTGGACCAACGCGTCT 420
Db 2691 CGGCTCCAGACCTATGACATGACATTTGGGACGCTGGAAGCTGGACCAACGCGTCT 2632

QY 421 CGTTGGCGTACGGAACGTGAATGTACGCGTGGGACGATACGTCGCGCTGTGAATC 480
Db 2631 CGTTGGCGTACGGAACGTGAATGTACGCGTGGGACGATACGTCGCGCTGTGAATC 2572

QY 481 ACTACGGAATAAAACGCTGGGACGTCACGTTCAACGCTACTCATGTTCAAGTGTGAC 540
Db 2571 ACTACGGAATAAAACGCTGGGACGTCACGTTCAACGCTACTCATGTTCAAGTGTGAC 2512

QY 541 GTGAGTGAAGAACATCCCGAATCCTTCGGAATGCGATGTTCAATCGGTGGAACCTTCCAA 600
Db 2511 GTGAGTGAAGAACATCCCGAATCCTTCGGAATGCGATGTTCAATCGGTGGAACCTTCCAA 2452

QY 601 CTGTAATCTCTGGGATTTGCTGGGACCTACGTAATTCGCGGTGCTCGATGTAAGAAAG 660
Db 2451 CTGTAATCTCTGGGATTTGCTGGGACCTACGTAATTCGCGGTGCTCGATGTAAGAAAG 2392

QY 661 ATGCGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGAAGAAATTTGATCTGGCGTT 720
Db 2391 ATGCGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGAAGAAATTTGATCTGGCGTT 2332

QY 721 GGTATGAGAGTTGTGTAGTGGCGACGAAAAAGTGTGACATGCGCAACCGTGCACCTGG 780
Db 2331 GGTATGAGAGTTGTGTAGTGGCGACGAAAAAGTGTGACATGCGCAACCGTGCACCTGG 2272

QY 781 GAATCCCTGTTGATGAGTGGCGGGATTCACAACGTTGGAACCTGTGGCGAATTT 840
Db 2271 GAATCCCTGTTGATGAGTGGCGGGATTCACAACGTTGGAACCTGTGGCGAATTT 2212

QY 841 GAGTTTGTTGATGAGAGCTCGGCAATGCGGTGCTCCGTGATTTGCTGCCGAGCT 900
Db 841 GAGTTTGTTGATGAGAGCTCGGCAATGCGGTGCTCCGTGATTTGCTGCCGAGCT 900

RESULT 3
AX548646/c 6334 bp DNA linear PAT 27-NOV-2002
LOCUS
DEFINITION
ACCESSION AX548646
VERSION AX548646.1 GI:25813614
KEYWORDS
SOURCE
ORGANISM
Rhodococcus erythropolis
Rhodococcus erythropolis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE
1. Blumci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
AUTHORS Rhodococcus cloning and expression vectors
TITLE Patent: WO 0205709-A 5 18-JUL-2002;
JOURNAL E.I. DU PONT DE NEMOURS' AND COMPANY (US)
FEATURES
source
1. .6334
/organism="Rhodococcus erythropolis"
/mol_type="genomic DNA"
/strain="AN12"
/db_xref="taxon:1833"
BATCH COUNT 1386 a 1807 c 1681 g 1460 t
ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 6334;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGTGAACAACCTTTCGCAAAAGACCGGCTCCCGTCTGATGTCG 60
Db 3051 ATGACGAGCGTAAGTGTGTGAACAACCTTTCGCAAAAGACCGGCTCCCGTCTGATGTCG 2992

QY 61 TCCGATTAAGCGCGCATCCGCGCAAGACCTGCGAACCCAACTTCAACAAATCACACGTC 120
Db 2991 TCCGATTAAGCGCGCATCCGCGCAAGACCTGCGAACCCAACTTCAACAAATCACACGTC 2932

QY 121 GAAACATTTAAAGCCCTGTGCGCGCGCATTTCTGGCGGTGAACGCTGTGACCATTTGTCAAC 180
Db 2931 GAAACATTTAAAGCCCTGTGCGCGCGCATTTCTGGCGGTGAACGCTGTGACCATTTGTCAAC 2872

QY 181 GGTCCGAAAGGTTCTGGATTCTGAGAGGCTTCTGTTCTCGCGAAAGGCTGGATCTGCCCC 240
Db 2871 GGTCCGAAAGGTTCTGGATTCTGAGAGGCTTCTGTTCTCGCGAAAGGCTGGATCTGCCCC 2812

QY 241 TGTGTGCGGAAAAGTGTGTGACATCTGTGACAGACGAAATTTCTCAAGTTGTGTCAT 300
Db 2811 TGTGTGCGGAAAAGTGTGTGACATCTGTGACAGACGAAATTTCTCAAGTTGTGTCAT 2752

QY 301 CAACCTGGGACTGGATCTGTTGGATGTGTGACATGACATGCGCCATACAGCTGGTCAG 360
Db 2751 CAACCTGGGACTGGATCTGTTGGATGTGTGACATGACATGCGCCATACAGCTGGTCAG 2692

QY 361 CGGCTCCAGACCTATGACATGACATTTGGGACGCTGGAAGCTGGACCAACGCGTCT 420
Db 361 CGGCTCCAGACCTATGACATGACATTTGGGACGCTGGAAGCTGGACCAACGCGTCT 420
```



QY 1021 TTGGGCGAGATCTTGGACCTGTCGACAGTGGCGGAGACTTGGGAAAATCTTCGTGATAC 1080  
Db 4032 TTGGCGGAGATCTTGGACCTGTCGACAGTGGCGGAGACTTGGGAAAATCTTCGTGATAC 3973  
QY 1081 TTGCATTAATGATGTCGCGGAGCGGATGTCGCGGCGGCGGATTAATGATGTCGCGGAGTGA 1140  
Db 3972 TTGCATTAATGATGTCGCGGAGCGGATGTCGCGGCGGCGGATTAATGATGTCGCGGAGTGA 3913

RESULT 5  
AX548647/c

LOCUS AX548647 11241 bp DNA linear PAT 27-NOV-2002  
DEFINITION Sequence 6 from Patent WO02055709.  
ACCESSION AX548647  
VERSION AX548647.1 GI:25813615  
KEYWORDS  
SOURCE Shuttle vector pRHR17  
ORGANISM Shuttle vector pRHR17  
REFERENCE 1  
artificial sequences; vectors.

REFERENCE  
AUTHORS Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.  
TITLE Rhodococcus cloning and expression vectors  
JOURNAL Patent: WO 02055709-A 6 18-JUL-2002  
E.I. DU PONT DE NEMOURS AND COMPANY (US)

FEATURES  
source  
1. 11241  
/organism="Shuttle vector pRHR17"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:214938"

BASE COUNT 2517 a 3109 c 2945 g 2670 t  
ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 11241;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACACGCGTGAAGTGTGTAACCTTTCCGCGAAGACCGGCTCTCGCTGCTGTCG 60  
Db 6641 ATGACACGCGTGAAGTGTGTAACCTTTCCGCGAAGACCGGCTCTCGCTGCTGTCG 6582  
QY 61 TTCCGATAGCGCGGATCTCGGACGACGACCTTCCCAAACTTCAACCAATCCACGCTCA 120  
Db 6581 TTCCGATAGCGCGGATCTCGGACGACGACCTTCCCAAACTTCAACCAATCCACGCTCA 6522  
QY 121 GAACATTTAAAGCGCTGTGCGCGCGCGGATTTCTGCGTGAAGCGGTGACCATTTGTCAAC 180  
Db 6521 GAACATTTAAAGCGCTGTGCGCGCGCGGATTTCTGCGTGAAGCGGTGACCATTTGTCAAC 6462  
QY 181 GGTCCGAAAGGTTCTGATTCGAGAGGCTTCTGCTGCGGAAAGGCTGATCTGCCCC 240  
Db 6461 GGTCCGAAAGGTTCTGATTCGAGAGGCTTCTGCTGCGGAAAGGCTGATCTGCCCC 6402  
QY 241 TGTCTGTGGGGGAAAGTGTGCGACATGTCGACGAAATTTCTCAAGTTGTCGTCAT 300  
Db 6401 TGTCTGTGGGGGAAAGTGTGCGACATGTCGACGAAATTTCTCAAGTTGTCGTCAT 6342  
QY 301 CAACCTCGGAGCTGATCTGTGCGATGTCGATGACATGACGCGCATACAGCTGTGAC 360  
Db 6341 CAACCTCGGAGCTGATCTGTGCGATGTCGATGACATGACGCGCATACAGCTGTGAC 6282  
QY 361 CGGCTCCACGACCTATGATGATGACCTTTCGAGAGCTTGGAAAAGCTGCGACCAAGGTCGT 420  
Db 6281 CGGCTCCACGACCTATGATGATGACCTTTCGAGAGCTTGGAAAAGCTGCGACCAAGGTCGT 6222  
QY 421 CGTTGGCGTACGGAACGTGAATGTAACGCTGCGACGATACGTCGCGGCTGTTAAATTC 480  
Db 6221 CGTTGGCGTACGGAACGTGAATGTAACGCTGCGACGATACGTCGCGGCTGTTAAATTC 6162  
QY 481 ACTCAAGGAAAAAGCGCTGACGCTGACGTTCAAGCGCTACTCATGTTCAAGTGTGAC 540  
Db 6161 ACTCAAGGAAAAAGCGCTGACGCTGACGTTCAAGCGCTACTCATGTTCAAGTGTGAC 6102

QY 541 GTGATGAGAAACATCTCGAATCCTTCTCGGATCCGATGTTTCAATCGGTGACCTTCCAAA 600  
Db 6101 GTGATGAGAAACATCTCGAATCCTTCTCGGATCCGATGTTTCAATCGGTGACCTTCCAAA 6042  
QY 601 CTCGTATCTCTGGGATTTGCTGCGCAGTACGTAATTCGGGTGCTCGATGTACGAAAG 660  
Db 6041 CTCGTATCTCTGGGATTTGCTGCGCAGTACGTAATTCGGGTGCTCGATGTACGAAAG 5982  
QY 661 ATCGGCGTGAAGCTGATCAAGTTCTGCTGCGATGTCGACGAAATTCGATCGCGTT 720  
Db 5981 ATCGGCGTGAAGCTGATCAAGTTCTGCTGCGATGTCGACGAAATTCGATCGCGTT 5922  
QY 721 GGTATGAGGTTGTGTAGTGTGCGGAGGAGGATCCAGAGCGTTGGAACCTGTGCGGAAATT 840  
Db 5921 GGTATGAGGTTGTGTAGTGTGCGGAGGAGGATCCAGAGCGTTGGAACCTGTGCGGAAATT 5862  
QY 781 GAATTCGCTGTTGATGACGTCGCGGAGGATCCAGAGCGTTGGAACCTGTGCGGAAATT 840  
Db 5861 GAATTCGCTGTTGATGACGTCGCGGAGGATCCAGAGCGTTGGAACCTGTGCGGAAATT 5802  
QY 841 GAGTTTGGTTCGATGAGGACGTCGCGGCAATGCGTGTCCGTCGATGTCGCGGAGCT 900  
Db 5801 GAGTTTGGTTCGATGAGGACGTCGCGGCAATGCGTGTCCGTCGATGTCGCGGAGCT 5742  
QY 901 GGTCTTGGGCGGAACTTAACGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTC 960  
Db 5741 GGTCTTGGGCGGAACTTAACGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTC 5682  
QY 961 ATGCTTGGGATCATTTCCGCGCGGATGTCGATGTCGATGTCGATGTCGATGTCGATGTC 1020  
Db 5681 ATGCTTGGGATCATTTCCGCGCGGATGTCGATGTCGATGTCGATGTCGATGTCGATGTC 5622  
QY 1021 TTGGCGGAGATCTTGGACCTGTCGAAAGCTGCGCGGACCTTGGGAAAATCTTCGTGATAC 1080  
Db 5621 TTGGCGGAGATCTTGGACCTGTCGAAAGCTGCGCGGACCTTGGGAAAATCTTCGTGATAC 5562  
QY 1081 TTGCATTAATGATGTCGCGGAGCGGATGTCGCGGCGGATTAATGATGTCGCGGAGTGA 1140  
Db 5561 TTGCATTAATGATGTCGCGGAGCGGATGTCGCGGCGGATTAATGATGTCGCGGAGTGA 5502

RESULT 6  
AY180162/c

LOCUS AY180162 11241 bp DNA circular SYN 01-MAY-2003  
DEFINITION Shuttle vector pRHR17, complete sequence.  
ACCESSION AY180162  
VERSION AY180162.1 GI:30313707  
KEYWORDS  
SOURCE Shuttle vector pRHR17  
ORGANISM Shuttle vector pRHR17  
REFERENCE 1  
artificial sequences; vectors.

REFERENCE  
AUTHORS Kostichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and  
Cheng, Q.

TITLE A small cryptic plasmid from Rhodococcus erythropolis:  
characterization and utility for gene expression

JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 11241)  
AUTHORS Kostichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and  
Cheng, Q.

TITLE Direct Submission  
JOURNAL Submitted (14-NOV-2002) CRØD, E. I. Dupont de Nemours Inc.,  
Experimental Station, Wilmington, DE 19880-0328, USA

FEATURES  
source  
1. 11241  
/organism="Shuttle vector pRHR17"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:214938"  
/focus

/note="Rhodococcus - Escherichia coli shuttle vector; Ssp  
I linearize pRHR12 cloned into the PvuII site of pBR328"  
join(9925..11241,1..3590)  
/organism="Cloning vector pBR328"

/mol\_type="genomic DNA"  
/db\_xref="taxon:221946"  
3591..9924  
/organism="Rhodococcus erythropolis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1833"  
/plasmid="PAN12"  
/note="cryptic plasmid"  
BASE COUNT 2517 a 3109 c 2945 g 2670 t  
ORIGIN

Query Match 100.0%; Score 1140; DB 12; Length 11241;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGACGAGCTTAAGTGTGACACCTTCCGGAAGACCGGCTCCGCTCTGCTGCG 60  
6641 ATGACGAGCTTAAGTGTGACACCTTCCGGAAGACCGGCTCCGCTCTGCTGCG 6582

61 TCCGATTAAGCGCGGATCCGACGACGACGACGACGACGACGACGACGACGAC 120  
6581 TCCGATTAAGCGCGGATCCGACGACGACGACGACGACGACGACGACGACGAC 6522

121 GAAACATTTAAGCGCTGTGCGCGGCGGATTTCTGCGGTGAACGGTGTGACAT 180  
6521 GAAACATTTAAGCGCTGTGCGCGGCGGATTTCTGCGGTGAACGGTGTGACAT 6462

181 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTCTGCTCTGCGGAAAGGCTGATG 240  
6461 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTCTGCTCTGCGGAAAGGCTGATG 6402

241 TCTGTGCGGAAAGTGTGACGACGACGACGACGACGACGACGACGACGACGAC 300  
6401 TCTGTGCGGAAAGTGTGACGACGACGACGACGACGACGACGACGACGACGAC 6342

301 CAATCGGAGCTGTGATCTGTGCGATGTGACGATGACGATGACGATGACGATG 360  
6341 CAATCGGAGCTGTGATCTGTGCGATGTGACGATGACGATGACGATGACGATG 6282

361 CGGCTCCGACGACCTTGTGACGATGACGATGACGATGACGATGACGATGACG 420  
6281 CGGCTCCGACGACCTTGTGACGATGACGATGACGATGACGATGACGATGACG 6222

421 CGTTGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 480  
6221 CGTTGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 6162

481 ACTCAACGAAAGGCTGTGACGACGACGACGACGACGACGACGACGACGACG 540  
6161 ACTCAACGAAAGGCTGTGACGACGACGACGACGACGACGACGACGACGACG 6102

541 GTGAGTGAAGACATCTCTGAAATCTTCTGGAATGGAATGGAATGGAATGGA 600  
6101 GTGAGTGAAGACATCTCTGAAATCTTCTGGAATGGAATGGAATGGAATGGA 6042

601 CTCGATCTCTGGAATTTGCTGCGACACGACGACGACGACGACGACGACGAC 660  
6041 CTCGATCTCTGGAATTTGCTGCGACACGACGACGACGACGACGACGACGAC 5982

661 ATCGGCGGTGAAGCTGATCAAGTCTCGTGGATCTGACGAAATTCGATCTGG 720  
5981 ATCGGCGGTGAAGCTGATCAAGTCTCGTGGATCTGACGAAATTCGATCTGG 5922

721 GGTATGAGAGTGTGATGAGTGAAGGACGAAAGGAGTGAAGGACGAAAGGAG 780  
5921 GGTATGAGAGTGTGATGAGTGAAGGACGAAAGGAGTGAAGGACGAAAGGAG 5862

781 GAAATGCTGTGATGACGATGAGGCGGAGATTCACAGCGTGAATCTGCGAGAT 840  
5861 GAAATGCTGTGATGACGATGAGGCGGAGATTCACAGCGTGAATCTGCGAGAT 5802

841 GAGTTGGTTGATGAGGAGCTGCGGCAATCGCGGTGCTCCGATGAGTGGTCCG 900

5801 GAGTTGGTTGATGAGGAGCTGCGGCAATCGCGGTGCTCCGATGAGTGGTCCG 5742

901 GGTCTTGGGCGAGAACTACAGATGCTCAGATGTTGAGCAGAGAAATCTGCCG 960

5741 GGTCTTGGGCGAGAACTACAGATGCTCAGATGTTGAGCAGAGAAATCTGCCG 5682

961 ATGTTTCCGATCATTTCCGCGCGCATCTGATATGATTTGAGCTTTGGCTTAC 1020

5681 ATGTTTCCGATCATTTCCGCGCGCATCTGATATGATTTGAGCTTTGGCTTAC 5622

1021 TTGCGGAGATCTCCGAGCTGTGGAAGCTGGCGGACCTTGGGAAATCTTGAT 1080

5621 TTGCGGAGATCTCCGAGCTGTGGAAGCTGGCGGACCTTGGGAAATCTTGAT 5562

1081 TTGATTTATGATTTCCCGGAGCGGATGTGCGGCGCGCGGATTAATATGAT 1140

5561 TTGATTTATGATTTCCCGGAGCGGATGTGCGGCGCGCGGATTAATATGAT 5502

RESULT 7  
AC133415/c  
LOCUS  
DEFINITION  
ACCESSION  
KEYWORDS  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

1 (bases 1 to 202897)  
Munzy,D,Marie, Metzger,M,Le, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alebrooks,S, Amin,A, Anguiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Beahmed,F, Bismail,K, Blair,C, Blankenburg,K, Blyth,P, Brown,M, Bryan,N, Buhay,C, Burch,P, Butrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Cessari,H, Centler,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,D, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, d'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denison,S, Deramo,C, Ding,Y, Dinh,H, Diya,K, Diaper,H, Dugan-Rocha,S, Dunn,A, Dudin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Fraser,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Georgegeorgis,B, Geier,K, Gill,R, Grady,M, Guerra,M, Guevara,M, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hilyk,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowals,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorensheva,L, Louisedge,H, Locado,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindartine,M, Mahmood,M, Mallory,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mawhinney,S, McLeod,M,P, McKell,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundaesa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwankwemeh,O, Okwomou,G, Olampunagoun,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankoch,C, Ploper,F, Poindeexter,A, Popovic,D, Primus,E, Pu,L, Puzo,M, Quiroz,J, Rachlin,B, Reeves,K, Regier,M,A, Reigh,R, Rives,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivas,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,W, Savery,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shwartsbeyn,A, Sisson,I, Sitter,C,D, Smaj,S, D, Snead,A, Sodergren,E, Song,X,Z, Sorelle,R, Soes,J, Steimle,M, Strong,R, Sutton,A, Swatek,A, Tabot,P, Taylor,C,

Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vales, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 202897)  
Rat Genome Sequencing Consortium.  
Submitted (12-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 202897)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:22795096.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with 'N's to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Project name: KBRK  
Center clone name: CH230-283C3  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 176284 bases at least Q40  
Consensus quality: 178546 bases at least Q30  
Consensus quality: 180221 bases at least Q20  
Estimated insert size: 183234; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 202897: contig of 202897 bp in length.  
Location/Qualifiers  
1..202897  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-283C3"  
2..752  
/note="clone boundary  
clone\_end:Sp6  
site:"

misc\_feature  
misc\_feature

end\_sequence:82256844"  
87041..88414  
/note="wgs contig"  
complement(198268..199155)  
/note="clone boundary  
clone\_end:77  
site:  
end\_sequence:82256842"  
200795..202897  
/note="wgs\_end\_extension  
clone\_end:77"  
misc\_feature  
BASE COUNT 45958 a 41690 c 43148 g 51167 t 20934 others  
ORIGIN  
Query Match 1.9%; Score 22; DB 2; Length 202897;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 765 CAACGTCGACCCCTGGAAATC 786  
Db 124096 CAACGTCGACCCCTGGAAATC 124075  
RESULT 8  
AC123219 224046 bp DNA linear HTG 22-SEP-2002  
LOCUS Rattus norvegicus clone CH230-175M19, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\* 6 unordered pieces.  
ACCESSION AC123219.3 GI:23195026  
VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 224046)  
REFERENCE  
AUTHORS Muzny, D.Marie, Metzker, M.Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Baves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, S., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, R., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewal, L., Louised, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, B., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Milosavljevic, A., Miner, G., Ming, A., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervill, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,





## TITLE

JOURNAL  
REFERENCE  
AUTHORS

## JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Mawhney, S., McLeod, M.P., McNeill, T.Z., Weenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaelelemeh, O., Okunolu, G., Olarunpadoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smag, S., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, U., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabot, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, S., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

## Direct Submission

Unpublished  
2 (bases 1 to 228864)

## Worley, K.C.

## Direct Submission

Submitted (25-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 228864)  
Rat Genome Sequencing Consortium.

## Direct Submission

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24942651. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

## Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Project Information

Center project name: GHTO

Center clone name: CH230-31N7

Summary Statistics

Assembly program: Atlas 3.0:

Consensus quality: 199066 bases at least Q40

Consensus quality: 203465 bases at least Q30

Consensus quality: 206920 bases at least Q20

Estimated insert size: 208252; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

NOTE: This is a 'working draft' sequence. It currently

consists of 6 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

## FEATURES

\* as soon as it is available and the accession number will  
\* be preserved.  
1 10379: contig of 10379 bp in length  
\* 10380 10479: gap of unknown length  
\* 10480 22035: contig of 21556 bp in length  
\* 22036 22213: gap of unknown length  
\* 22213 22293: contig of 1158 bp in length  
\* 22293 22394: gap of unknown length  
\* 22394 22512: contig of 1727 bp in length  
\* 22512 22520: gap of unknown length  
\* 22520 22645: contig of 1237 bp in length  
\* 22645 22655: gap of unknown length  
\* 22655 22864: contig of 2307 bp in length.  
Location/Qualifiers  
1..228864  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-31N7"  
1..1039  
/note="wgs end extension  
clone end:Sp6"  
complement(2260..3078)  
/note="clones boundary  
clone end:Sp6  
site:BCORI  
end sequence: BH324953"  
10480..12201  
/note="wgs contig"  
41569..42701  
/note="wgs contig"  
98777..101307  
/note="wgs contig"

BASE COUNT 56247 a 49203 c 49836 g 54580 t 18998 others

Query Match 1.9%; Score 22; DB 2; Length 228864;

Best local similarity 100.0%; Pred. No. 1.9;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OR 765 CAACCGTCGACCGTCGGAATC 786

Db 202492 CAACCGTCGACCGTCGGAATC 202513

## RESULT 10

## LOCUS

AP005161 141405 bp DNA linear HTG 17-MAY-2002

Oryza sativa (japonica cultivar-group) chromosome 8 clone

OSJNBa0036E18, \*\*\* SEQUENCING IN PROGRESS \*\*\*.

AP005161 GI:20975287

HTG: HTGS PHASR2

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

## REFERENCE

## AUTHORS

Sasaki, T., Matsumoto, T. and Katayose, Y.

Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC

clone: OSJNBa0036E18

Published Only in Database (2002)

## JOURNAL

## AUTHORS

Sasaki, T., Matsumoto, T. and Katayose, Y.

Submitted (15-MAY-2002) Takuji Sasaki, National Institute of

Agrrobiological Sciences, Rice Genome Research Program; Kannondai

2-1-2, Tsukuba, Ibaraki 305-8602, Japan

(E-mail: tsasaki@nars.affrc.go.jp, URL: <http://rsgp.dna.affrc.go.jp/>,  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

NOTE: It currently consists of 1 contigs. Gaps between the contigs

are represented as runs of N. The order of the pieces is believed



to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

\* NOTE: This is a 'working draft' sequence.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

## FEATURES

Location/Qualifiers

1. 141405

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="8"

/clone="OSJNBa0036R18"

BASE COUNT 41811 a 28829 c 29018 g 41609 t 138 others

## ORIGIN

Query Match 1.8%; Score 21; DB 2; Length 141405;

Best Local Similarity 100.0%; Pred. No. 7.4;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1054 GCGACTTGCGAAATCTTCGT 1074

Db 22729 GCGACTTGCGAAATCTTCGT 22749

RESULT 11  
AP005799 155547 bp DNA linear HTG 03-OCT-2002

LOCUS Oryza sativa (japonica cultivar-group) chromosome 8 clone

DEFINITION OSJNBa0012K14, \*\*\* SEQUENCING IN PROGRESS \*\*\*.

ACCESSION AP005799

VERSION AP005799.1 GI:23496503

KEYWORDS HTG; HTGS\_PHASE2.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Eriactoidae; Oryzaceae; Oryza.

## REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

## ORIGIN

Query Match 1.8%; Score 21; DB 2; Length 155547;

Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1054 GCGACTTGCGAAATCTTCGT 1074

Db 147013 GCGACTTGCGAAATCTTCGT 147033

## RESULT 12

AX548660/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

## ORIGIN

Query Match 1.8%; Score 20; DB 6; Length 167;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 GATTAATATCGTTCCGACT 1138

Db 20 GATTAATATCGTTCCGACT 1

## RESULT 13

AX049359/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

## ORIGIN

Query Match 1.8%; Score 20; DB 6; Length 167;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 GATTAATATCGTTCCGACT 1138

Db 20 GATTAATATCGTTCCGACT 1

## RESULT 13

AX049359/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

## ORIGIN

Query Match 1.8%; Score 20; DB 6; Length 167;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 GATTAATATCGTTCCGACT 1138

Db 20 GATTAATATCGTTCCGACT 1

## RESULT 13

AX049359/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

## ORIGIN

Query Match 1.8%; Score 20; DB 6; Length 167;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 GATTAATATCGTTCCGACT 1138

Db 20 GATTAATATCGTTCCGACT 1

## RESULT 13

AX049359/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

## ORIGIN

Query Match 1.8%; Score 20; DB 6; Length 167;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1119 GATTAATATCGTTCCGACT 1138

Db 20 GATTAATATCGTTCCGACT 1

## RESULT 13

AX049359/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

QY 324 GATGTCAGATGACCATGC 343  
Db 81 GATGTCAGATGACCATGC 62

RESULT 14  
LOCUS AY227111/c 1402 bp mRNA linear PRI 12-FEB-2003  
DEFINITION Homo sapiens clone IMAGE:2549054 truncated transmembrane transport protein (SLC36A1) mRNA, complete cds; alternatively spliced.  
ACCESSION AY227111  
VERSION AY227111.1 GI:28372397  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1402)  
AUTHORS Bermingham, J.R. Jr. and Pennington, J.  
TITLE Genome organization and expression of the SLC36a cluster of amino acid transport protein genes  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1402)  
AUTHORS Bermingham, J.R. Jr. and Pennington, J.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-2003) McLaughlin Research Institute, 1520 23rd Street South, Great Falls, MT 59405, USA  
Location/Qualifiers  
1..1420  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q33"  
/clone="IMAGE:2549054"  
1..1402  
/gene="SLC36A1"  
/note="LYAAT-1"  
92..95  
/gene="SLC36A1"  
/note="apparently deleted and replaced by a 62 nt sequence in the genomic sequence"  
/replace="  
144..1304  
/gene="SLC36A1"  
/note="alternatively spliced; uses alternative 3' exon that results in C-terminal truncation; transmembrane protein"  
/codon\_start=1  
/product="truncated transmembrane transport protein"  
/protein\_id="AAO37091.1"  
/db\_xref="GI:28372398"  
/translation="MSTQRLRNEDYHDSYDVSPEPSSEGLNNLSPGSYQRFQSS NSTWPTLIHLKGNIGIGLGLPLAVNAGIVMGISLIIITIGYAVACMGIVYCA HHRCRLINKSPVDYGDPTWYVGGSSCSMLRHHAKGRVNDPFLIVYLOGCCVYFA FLADNFKQVTEANGTNNCHNNEVILPTPMDSRLYMLSPLEFLVLFIRNRLAS IFSLLANTMLVSLVMIYQFIVQRIIPDSHLPLVAVKTYPLFFGTAISFEGIGVLL PLENKMKDPRKRPPLIYLGMVITVILYISGLCYGFQFGANIQGSITLNDPNCMLYQS VKLYVSIIGIFFTYALQFYVPAKIIIPFFVSRAPEHCELVDLFEVRLVCLT"

BASE COUNT 312 a 393 c 352 g 345 t

ORIGIN

Query Match 1.8%; Score 20; DB 9; Length 1402;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTCAGATGACCATGC 343  
Db 1046 GATGTCAGATGACCATGC 1027

RESULT 15  
LOCUS AY227112/c 1420 bp mRNA linear PRI 12-FEB-2003

DEFINITION Homo sapiens clone DKFZp434G1123 transmembrane transport protein (SLC36A1) mRNA, partial cds.  
ACCESSION AY227112  
VERSION AY227112.1 GI:28372399  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1420)  
AUTHORS Bermingham, J.R. Jr. and Pennington, J.  
TITLE Genome organization and expression of the SLC36a cluster of amino acid transport protein genes  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1420)  
AUTHORS Bermingham, J.R. Jr. and Pennington, J.  
TITLE Direct Submission  
JOURNAL Submitted (28-JAN-2003) McLaughlin Research Institute, 1520 23rd Street South, Great Falls, MT 59405, USA  
Location/Qualifiers  
1..1420  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q33"  
/clone="DKFZp434G1123"  
/note="derived from dbEST AL043182"  
1..1420  
/gene="SLC36A1"  
/note="LYAAT-1"  
1..942  
/gene="SLC36A1"  
/note="transmembrane transport protein"  
/product="transmembrane transport protein"  
/protein\_id="AAO37092.1"  
/db\_xref="GI:28372400"  
/translation="DNFKQVTEANGTNNCHNNEVILPTPMDSRLYMLSPLEFLVL LVFIRNRLASIFSLLANTMLVSLVMIYQFIVQRIIPDSHLPLVAVKTYPLFFGTA IFSFEGIGVLLPLENKMKDPKRPPLIYLGMVITVILYISGLCYGFQFGANIQGSIT LNPNCMLYQS VKLYVSIIGIFFTYALQFYVPAKIIIPFFVSRAPEHCELVDLFEVRLV CLTCITLAILPRLDVLSLGSVSSSLALIIPLLEVTTFYSBGMSPLTIFKDAL ISILGFVGVVGTVEALYELLIQPSNAPIFINSTCAFI"

BASE COUNT 281 a 456 c 315 g 368 t

ORIGIN

Query Match 1.8%; Score 20; DB 9; Length 1420;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTCAGATGACCATGC 343  
Db 414 GATGTCAGATGACCATGC 395

Search completed: November 8, 2003, 22:29:30  
Job time : 2941 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 20:05:57 ; Search time 281 Seconds  
(without alignments)

10951.457 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1140  
Sequence: 1 atgaccagcgcgaagtcgtga.....taatatcggttcgcaagtcga 1140

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2552756 segs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1140	100.0	1140	24	ABQ76122
2	1140	100.0	6334	24	ABQ76124
3	1140	100.0	9652	24	ABQ76126
4	1140	100.0	11241	24	ABQ76125
5	20	1.8	20	24	ABQ76138
6	20	1.8	167	23	AB157927
7	20	1.8	739	23	AAS90653
8	20	1.8	947	23	ABV23244

C 9	20	1.8	947	23	ABV29088	Human prostate exp
C 10	20	1.8	1428	23	ABL57930	Human VG51 coding
C 11	20	1.8	1501	25	ABX70643	Human cDNA encoding
C 12	20	1.8	1513	25	ABX70642	Human cDNA encoding
C 13	20	1.8	1597	25	ABX70783	Novel human cDNA s
C 14	20	1.8	1771	23	AAS90655	DNA encoding novel
C 15	20	1.8	2057	25	ABZ80236	Human ttramodrin 3
C 16	20	1.8	2807	25	ABX12019	Human transporter and 1
C 17	20	1.8	3114	24	AA144674	Human transporter
C 18	20	1.8	249999	25	ABZ80229	Human ttramodrin ge
C 19	19	1.7	332	21	AAC72811	Single nucleotide
C 20	19	1.7	476	25	ABX22328	Human GDP-mannose
C 21	19	1.7	1209	24	ABV78077	Hypoxia-regulated
C 22	19	1.7	1333	22	AAH34705	Human colon cancer
C 23	19	1.7	3810	24	ABN97214	Gene #3712 used to
C 24	19	1.7	4515	23	ABL07083	Drosophila melanog
C 25	19	1.7	4578	10	AAN92192	cDNA sequence of c
C 26	19	1.7	4579	11	AAQ06287	L-A virus genome.
C 27	19	1.7	4580	19	AAV05284	Saccharomyces cere
C 28	19	1.7	26370	23	ABL07082	Rhodococcus AN12 d
C 29	18	1.6	20	24	ABQ76139	Rhodococcus AN12 d
C 30	18	1.6	282	25	ABX25710	Human GDP-mannose
C 31	18	1.6	399	22	AAS35882	Human cardiovascular
C 32	18	1.6	440	22	AAS35179	Human cardiovascular
C 33	18	1.6	608	21	AAQ07739	Fusarium venenatum
C 34	18	1.6	619	23	AAV73050	DNA encoding novel
C 35	18	1.6	807	17	AAV15535	DNA encoding rat c
C 36	18	1.6	982	19	AAV52040	Helicobacter polyP
C 37	18	1.6	1065	23	AAV71237	DNA encoding novel
C 38	18	1.6	1095	23	AAS52556	E. coli DNA for ce
C 39	18	1.6	1659	23	AAS82835	DNA encoding novel
C 40	18	1.6	4104	24	ABX35783	Human IGF1R polyru
C 41	18	1.6	4104	24	ABX10026	Human IGF1R DNA fr
C 42	18	1.6	4104	24	ABV78207	Human IGF1R DNA SE
C 43	18	1.6	4104	24	AB191748	Human polynucleoti
C 44	18	1.6	4383	25	ABZ71068	Mycobacterium tube
C 45	18	1.6	4575	15	AAQ73445	IGF-I receptor. H

#### ALIGNMENTS

RESULT 1  
ABQ76122  
ID ABQ76122 standard; DNA; 1140 BP.  
XX  
AC ABQ76122;  
XX  
DT 13-JUN-2003 (first entry)  
XX  
DE Rhodococcus AN12 replication protein Rep DNA.  
XX  
KW Plasmid stability protein; replication protein; ethylene forming enzyme;  
KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
KW polyhydroxyalkanoic acid synthase; PBA synthase; nitrile hydratase;  
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase; gene;  
KW shuttle vector; Rep; ds.  
XX  
OS Rhodococcus erythropolis.  
XX  
FN ~~MO200255709-A2.~~  
XX  
PD 18-JUL-2002.  
XX  
PF 12-DEC-2001; 2001MO-US47868.  
XX  
PR 12-DEC-2000; 2000US-254868P.  
XX  
PA (DUPO) DU FONT DE NEMOURS & CO E. I.  
XX  
PI ~~Branucci MG.~~ Cheng Q, Kostichka KN, Tomb J,  
XX  
DR WFI, 2002-557827/59.

DR P-PSDB; ABB84278.

XX New nucleic acid molecule encoding replication protein/plasmid

PT stability protein, useful in cloning and expression vectors,

PT particularly shuttle vectors for expression of heterologous genes in

PT Rhodococcus species -

XX

PS Claim 2; Page 63-64; 96pp; English.

XX

CC This invention describes a novel nucleic acid encoding a replication

CC protein or a plasmid stability protein. The product of the invention is

CC useful for expression of nucleic acid such as genes encoding enzymes

CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic

CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile

CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol

CC dehydrogenase, terpene synthases, and cholesterol oxidase in an

CC Actinomycetaceae bacteria. The replication protein or plasmid stability

CC protein are useful in cloning and expression vectors and particularly in

CC shuttle vectors for the expression of homologous and heterologous genes

CC in Rhodococcus sp. This sequence encodes the Rhodococcus AN12 Rep

CC protein described in the disclosure of the invention.

XX

SQ Sequence 1140 BP; 241 A; 287 C; 347 G; 265 T; 0 other;

Query Match 100.0%; Score 1140; DB 24; Length 1140;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCGAGGTGATGCTGTAACACCTTTCCGCGAAAGACCGGCTCTCCCTCGTGTG 60

DB 1 ATGACCGAGGTGATGCTGTAACACCTTTCCGCGAAAGACCGGCTCTCCCTCGTGTG 60

QY 61 TTCGATTAAGCGCGGCGATCCGCGACGAACTGCGAACCTTCAAAATCAACAGTCA 120

DB 61 TTCGATTAAGCGCGGCGATCCGCGACGAACTGCGAACCTTCAAAATCAACAGTCA 120

QY 121 GAAACATTTAAGCGCGGCGATCCGCGACGAACTGCGAACCTTCAAAATCAACAGTCA 180

DB 121 GAAACATTTAAGCGCGGCGATCCGCGACGAACTGCGAACCTTCAAAATCAACAGTCA 180

QY 181 GGTCCGAAAGGTTCGTGATTCGGAAGCCCTTCTCTGCGGAAAGGCTGTGATCTGCCC 240

DB 181 GGTCCGAAAGGTTCGTGATTCGGAAGCCCTTCTCTGCGGAAAGGCTGTGATCTGCCC 240

QY 241 TGTCTGCGGGAAGGTTCGTGATTCGGAAGCCCTTCTCTGCGGAAAGGCTGTGATCTGCCC 300

DB 241 TGTCTGCGGGAAGGTTCGTGATTCGGAAGCCCTTCTCTGCGGAAAGGCTGTGATCTGCCC 300

QY 301 CAACCTGCGGAGCTGATCTGTTGCGATGATGATGATGATGATGATGATGATGATGATGAT 360

DB 301 CAACCTGCGGAGCTGATCTGTTGCGATGATGATGATGATGATGATGATGATGATGATGAT 360

QY 361 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420

DB 361 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420

QY 421 CGTTGGGCTACGGAAGCTGGAATGTAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAG 480

DB 421 CGTTGGGCTACGGAAGCTGGAATGTAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAG 480

QY 481 ACTACGGAAGAAAGGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAG 540

DB 481 ACTACGGAAGAAAGGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAG 540

QY 541 GTGAGTGAAGAACATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

DB 541 GTGAGTGAAGAACATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY 601 CTGCTATCTCTGGGATTTGCTGCGGCACTGATGATGATGATGATGATGATGATGATGATGATGAT 660

DB 601 CTGCTATCTCTGGGATTTGCTGCGGCACTGATGATGATGATGATGATGATGATGATGATGATGAT 660

QY 661 ATGCGCGTGAAGCTGATCAAGTCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 720

DB 661 ATGCGCGTGAAGCTGATCAAGTCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 720

QY 721 GGTATGAGAGTTGGTATGAGCTGCGACGGAAGAAAGTGTGATGATGATGATGATGATGATGATGAT 780

DB 721 GGTATGAGAGTTGGTATGAGCTGCGACGGAAGAAAGTGTGATGATGATGATGATGATGATGATGAT 780

QY 781 GAAATCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840

DB 781 GAAATCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840

QY 841 GAGTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900

DB 841 GAGTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900

QY 901 GGTCTGGGCGAGAACTAACAGATGCTCAAGTCTGTTGAGCAGAAAGATCTGCCCGGCTC 960

DB 901 GGTCTGGGCGAGAACTAACAGATGCTCAAGTCTGTTGAGCAGAAAGATCTGCCCGGCTC 960

QY 961 ATGCTTGGGCGAGAACTAACAGATGCTCAAGTCTGTTGAGCAGAAAGATCTGCCCGGCTC 1020

DB 961 ATGCTTGGGCGAGAACTAACAGATGCTCAAGTCTGTTGAGCAGAAAGATCTGCCCGGCTC 1020

QY 1021 TTCGCGGAGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080

DB 1021 TTCGCGGAGATCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080

QY 1081 TTGCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140

DB 1081 TTGCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140

RESULT 2

AB076124/c

ID AB076124 standard; DNA; 6334 BP.

XX

AC AB076124;

XX

DT 13-JAN-2003 (first entry)

XX

DE Rhodococcus AN12 derived plasmid pAN12 DNA.

XX

KW Plasmid stability protein; replication protein; ethylene forming enzyme;

KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;

KW polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;

KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;

KW shuttle vector; circular; ds.

XX

OS Rhodococcus erythropolis.

OS Synthetic.

XX

PN WO200255709-A2.

XX

PD 18-JUL-2002.

XX

PF 12-DEC-2001; 2001WO-US47868.

XX

PR 12-DEC-2000; 2000US-254868P.

XX

PA (DUPO ) DU POINT DE NEMOURS & CO E I.

XX

PI **BzAmuicf MG, Cheng Q, Kostichka KN, Tomb J;**

PI **BzAmuicf MG, Cheng Q, Kostichka KN, Tomb J;**

XX

DR WPI; 2002-557827/59.

XX

XX New nucleic acid molecule encoding replication protein/plasmid

PT stability protein, useful in cloning and expression vectors,

PT particularly shuttle vectors for expression of heterologous genes in

PT Rhodococcus species -

PS Claim 19; Page 68-71; 96pp; English.

XX

CC This invention describes a novel nucleic acid encoding a replication

CC protein or a plasmid stability protein. The product of the invention is  
CC useful for expression of nucleic acid such as genes encoding enzymes  
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
CC actinomycetale bacteria. The replication protein or plasmid stability  
CC protein are useful in cloning and expression vectors and particularly in  
CC shuttle vectors for the expression of homologous and heterologous genes  
CC in Rhodococcus sp. This sequence represents the Rhodococcus AN12 derived  
CC plasmid pAN12 DNA described in the disclosure of the invention.

XX Sequence 6334 BP; 1386 A; 1807 C; 1681 G; 1460 T; 0 other:

Query Match 100.0%; Score 1140; DB 24; Length 6334;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCAAGCTTAAGTGTGAACACCTTCCGGCAAGACCGGCTCCCGCTCTGTGTG 60  
DB 3051 ATGACCAAGCTTAAGTGTGAACACCTTCCGGCAAGACCGGCTCCCGCTCTGTGTG 2992  
QY 61 TTCGATTAAGCGCGGCAATCCGGGCAAGATCGAAGCCCAACTTGAACCAATCACCACTCA 120  
DB 2991 TTCGATTAAGCGCGGCAATCCGGGCAAGATCGAAGCCCAACTTGAACCAATCACCACTCA 2932  
QY 121 GAAACATTTTAACGCTGTGCGCGGCGGATTTCTGGCTGGAACGGTGTACATTGTCAAC 180  
DB 2931 GAAACATTTTAACGCTGTGCGCGGCGGATTTCTGGCTGGAACGGTGTGTCAAC 2872  
QY 181 GGTCCGAAGGTTCTGATTCGAGGCGCTTCTTCCGCGAAGGCGCTGATTCGCC 240  
DB 2871 GGTCCGAAGGTTCTGATTCGAGGCGCTTCTTCCGCGAAGGCGCTGATTCGCC 2812  
QY 241 TGCCTGTCGGGAAAGTCGTCGCAATCGTCGCAAGAAATTTCTCAAGTTGTGTGCAT 300  
DB 2811 TGCCTGTCGGGAAAGTCGTCGCAATCGTCGCAAGAAATTTCTCAAGTTGTGTGCAT 2752  
QY 301 CAACTCGGGAATGATCTGTTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
DB 2751 CAACTCGGGAATGATCTGTTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 2692  
QY 361 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
DB 2691 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2632  
QY 421 CGTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
DB 2631 CGTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2572  
QY 481 ACTACGGAAGAAACGCTGCGACGTCACGTTCCAGGCTACTACTACTACTACTACTACTACTACT 540  
DB 2571 ACTACGGAAGAAACGCTGCGACGTCACGTTCCAGGCTACTACTACTACTACTACTACTACTACT 2512  
QY 541 GTGAGTGAAGCAATCTCGAATCTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 2511 GTGAGTGAAGCAATCTCGAATCTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 2452  
QY 601 CTGATATCTTGGGATTTGCTGCGGCTACTAGTATTTGGGCTGCTGATGATGATGATGATGAT 660  
DB 2451 CTGATATCTTGGGATTTGCTGCGGCTACTAGTATTTGGGCTGCTGATGATGATGATGATGAT 2392  
QY 661 ATCGGCGGTGAAGCTGATCAAGTTCTCGTGTGCTGATGATGATGATGATGATGATGATGATGAT 720  
DB 2391 ATCGGCGGTGAAGCTGATCAAGTTCTCGTGTGCTGATGATGATGATGATGATGATGATGATGAT 2332  
QY 721 GGTATGAGAGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
DB 2331 GGTATGAGAGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2272  
QY 781 GAAATCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
DB 2271 GAAATCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2212

QY 841 GAGTTTGGTTCAGTGGGACGTCGGGCAATCGGCTGATCCCGTGAATTCGCTCCCGAGCT 900  
DB 2211 GAGTTTGGTTCAGTGGGACGTCGGGCAATCGGCTGATCCCGTGAATTCGCTCCCGAGCT 2152  
QY 901 GGTCTTGGGCGAAGCTTAACAGATGCTCAGATGTTGAGCAGAGAAATCTCCCGGCTC 960  
DB 2151 GGTCTTGGGCGAAGCTTAACAGATGCTCAGATGTTGAGCAGAGAAATCTCCCGGCTC 2092  
QY 961 ATGCTTGGCATATTCGCGCGGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
DB 2091 ATGCTTGGCATATTCGCGCGGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2032  
QY 1021 TTCGGGAGATTCCTGCGGCTGCTGCGAAGCTGCGGCGGCTTGGGAAATCTTGGTATGAC 1080  
DB 2031 TTCGGGAGATTCCTGCGGCTGCTGCGAAGCTGCGGCGGCTTGGGAAATCTTGGTATGAC 1972  
QY 1081 TTGCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
DB 1971 TTGCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1912

### RESULT 3

AB076126/c  
ID AB076126 standard; DNA; 9652 BP.

XX AB076126;

DT 13-JAN-2003 (first entry)

XX Plasmid pRHB171 DNA.

KM Plasmid stability protein; replication protein; ethylene forming enzyme;  
KM carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
KM polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;  
KM alcohol dehydrogenase; terpene synthase; cholesterol oxidase;  
KM shuttle vector; circular; ds.

XX Synthetic.

OS WO200255709-A2.

PN 18-JUL-2002.

PD 12-DEC-2001; 2001WO-US47868.

PR 12-DEC-2000; 2000US-254868P.

XX (DUPO) DU PONT DE NEMOURS & CO E. I.

XX **Bramucci MG, Cheng Q, Kostichka KN, Tomb J;**

XX **WPI; 2002-557827/59.**

PT New nucleic acid molecule encoding replication protein/plasmid  
PT stability protein, useful in cloning and expression vectors,  
PT particularly shuttle vectors for expression of heterologous genes in  
PT Rhodococcus species

PS Claim 27; Page 72; 96pp; English.

CC This invention describes a novel nucleic acid encoding a replication  
CC protein or a plasmid stability protein. The product of the invention is  
CC useful for expression of nucleic acid such as genes encoding enzymes  
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
CC actinomycetale bacteria. The replication protein or plasmid stability  
CC protein are useful in cloning and expression vectors and particularly in  
CC shuttle vectors for the expression of homologous and heterologous genes  
CC in Rhodococcus sp. This sequence represents the plasmid pRHB171 DNA  
CC described in the disclosure of the invention.

XX Sequence 9652 BP; 2097 A; 2729 C; 2559 G; 2267 T; 0 other;  
SQ  
Query Match 100.0%; Score 1140; DB 24; Length 9652;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGTAACACCTTTCCGCAAGAACCGGCGCTCCGCTCGTGTG 60  
DB 5052 ATGACGAGCGTAAGTGTGTAACACCTTTCCGCAAGAACCGGCGCTCCGCTCGTGTG 4993

QY 61 TCCGATTAAGCGCGGCAATCCGCAAGAACTGCGAACCAATTCAACAAATCACCAGTCA 120  
DB 4992 TCCGATTAAGCGCGGCAATCCGCAAGAACTGCGAACCAATTCAACAAATCACCAGTCA 4933

QY 121 GAAACATTAAACCCCTGTGCGCGCGATTTCTGGGTGTAACGGTGTGACATTGTCAAC 180  
DB 4932 GAAACATTAAACCCCTGTGCGCGCGATTTCTGGGTGTAACGGTGTGACATTGTCAAC 4873

QY 181 GGTCCGAAAGGTTCTGGAATCCGAGGCGCTTCTTCCGCGGAAAGGCGTGGATCTGCCCC 240  
DB 4872 GGTCCGAAAGGTTCTGGAATCCGAGGCGCTTCTTCCGCGGAAAGGCGTGGATCTGCCCC 4813

QY 241 TGTGTGCGGGAAGGTCGATGCAATGTCGACAGCAAAATTTCTCAAGTTGTTGCTCAT 300  
DB 4812 TGTGTGCGGGAAGGTCGATGCAATGTCGACAGCAAAATTTCTCAAGTTGTTGCTCAT 4753

QY 301 CAATCGGGAATGGAATGTTGGGATGTAACGATGCAATGCGCCATACAGCTGGTCA 360  
DB 4752 CAATCGGGAATGGAATGTTGGGATGTAACGATGCAATGCGCCATACAGCTGGTCA 4693

QY 361 CGGCTCCAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
DB 4692 CGGCTCCAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4633

QY 421 CGTTGCGGTAACGGAAGTGAATGTAACGCTGCGAGCAATGCTGCGCGCTGTTGAATC 480  
DB 4632 CGTTGCGGTAACGGAAGTGAATGTAACGCTGCGAGCAATGCTGCGCGCTGTTGAATC 4573

QY 481 ACTACGGAAGAAACCGCTGCGACGTCACGTTACCGCTACATGTTCAAGTGAAC 540  
DB 4572 ACTACGGAAGAAACCGCTGCGACGTCACGTTACCGCTACATGTTCAAGTGAAC 4513

QY 541 GTGAGTGAAGACATCTCGAATCCTTCTCGGATCGAATGTTGATCGGTGACCTTCCAAA 600  
DB 4512 GTGAGTGAAGACATCTCGAATCCTTCTCGGATCGAATGTTGATCGGTGACCTTCCAAA 4453

QY 601 CTGCTATCTCGGATTTGCTGCGCACCTACGTAATTCGGTGTCTCGATGTACGAAG 660  
DB 4452 CTGCTATCTCGGATTTGCTGCGCACCTACGTAATTCGGTGTCTCGATGTACGAAG 4393

QY 661 ATCGGCGGTGAAGCGTCAATCAAGTTCTGCTGCGTATCGACGAAATTTGATCGCGTT 720  
DB 4392 ATCGGCGGTGAAGCGTCAATCAAGTTCTGCTGCGTATCGACGAAATTTGATCGCGTT 4333

QY 721 GGTATGAGGTTGTAAGTGCAGCGAAGAAAGTGTGCAATGCGCAACCGTGCACCTCTGG 780  
DB 4332 GGTATGAGGTTGTAAGTGCAGCGAAGAAAGTGTGCAATGCGCAACCGTGCACCTCTGG 4273

QY 781 GAAATCGCTGTTGATGCAATGCGCGGGAATCCACAGGCTTTGAACTGTGCGCAAGATT 840  
DB 4272 GAAATCGCTGTTGATGCAATGCGCGGGAATCCACAGGCTTTGAACTGTGCGCAAGATT 4213

QY 841 GAGTTGGTTCGATGAGGAGCGTCCGCGCAATCGCGTGGTCTCCGATGATGCGGAGCT 900  
DB 4212 GAGTTGGTTCGATGAGGAGCGTCCGCGCAATCGCGTGGTCTCCGATGATGCGGAGCT 4153

QY 901 GGTCTTGGGCGAAGCACTAAGATCTGATGTTGACAGGAAATCTGCCCCGCTC 960  
DB 4152 GGTCTTGGGCGAAGCACTAAGATCTGATGTTGACAGGAAATCTGCCCCGCTC 4093

QY 961 ATGTTGCGATCATTCGCGCGGATCGGTGATGATGATTCGACTTGTGCGCTTAACCTC 1020

DB 4092 ATGTTGCGATCATTCGCGCGGATCGGTGATGATGATTCGGAATTTGCGCTTAACCTC 4033

QY 1021 TTCCGCGAAGATCTCGAATCTGTCGAGCTGCGCGGCACTTGGGAAATCTTGTGATCAC 1080

DB 4032 TTCCGCGAAGATCTCGAATCTGTCGAGCTGCGCGGCACTTGGGAAATCTTGTGATCAC 3973

QY 1081 TTGCAATTATCGATTTCCCGCGACGGAATGTGCGGCCCCCGGATTAATTCGTTGCAAGTGA 1140

DB 3972 TTGCAATTATCGATTTCCCGCGACGGAATGTGCGGCCCCCGGATTAATTCGTTGCAAGTGA 3913

## RESULT 4

AB076125/c  
ID AB076125 standard; DNA; 11241 BP.

AB076125;

13-JUN-2003 (first entry)

Plasmid pRHR17 DNA.

Plasmid stability protein; replication protein; ethylene forming enzyme;

carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;

polyhydroxyalkanoic acid synthase; pHA synthase; nitrile hydratase;

alcohol dehydrogenase; terpene synthase; cholesterol oxidase;

shuttle vector; circular; ds.

Synthetic.

WO200255709-A2.

18-JUL-2002.

12-DEC-2001; 2001WO-US47868.

12-DEC-2000; 2000US-254868P.

(DUPO) DU PONT DE NEMOURS & CO E. I.

Brämmuccl MG, Cheng Q, Kostichka KN, Tomb J;

WPI; 2002-557827/59.

Claim 26; Page 71-72; 96pp; English.

This invention describes a novel nucleic acid encoding a replication

protein or a plasmid stability protein. The product of the invention is

involved in the production of isoprenoid molecules, polyhydroxyalkanoic

acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile

hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol

dehydrogenase, terpene synthases, and cholesterol oxidase in an

Actinomyetales bacteria. The replication protein or plasmid stability

shuttle vectors are useful in cloning and expression vectors and particularly in

in Rhodococcus sp. This sequence represents the plasmid pRHR17 DNA

described in the disclosure of the invention.

Sequence 11241 BP; 2517 A; 3109 C; 2945 G; 2670 T; 0 other;

Query Match 100.0%; Score 1140; DB 24; Length 11241;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGTAACACCTTTCCGCAAGAACCGGCGCTCCGCTCGTGTG 60  
DB 6641 ATGACGAGCGTAAGTGTGTAACACCTTTCCGCAAGAACCGGCGCTCCGCTCGTGTG 6582





DE Human VG51 exon 8.  
 XX Antiaesthetic; anxiolytic; antiepileptic; antihypertensive; human;  
 XX psychotropic; glutamate transporter; transporter; GABA;  
 KM gamma-aminobutyric acid transporter; GABA transporter; neurotransmitter;  
 KM asthma; anxiety; epilepsy; hypertension; psychiatric disorder;  
 KM neurotic disorder; VG51; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200071709-A1.  
 XX  
 PD 30-NOV-2000.  
 XX  
 PF 19-MAY-2000; 2000WO-FR01383.  
 XX  
 PR 21-MAY-1999; 99FR-0006525.  
 XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Gires B, Gasnier B, Sagne C, El Mestikawy S, Hamon M;  
 XX WPI; 2001-025160/03.  
 DR  
 XX New mammalian amino acid transporter, used e.g. to screen for  
 PT psychotropic agents, is high capacity but low affinity transporter of  
 PT gamma-aminobutyric acid -  
 XX  
 PS Claim 5; Page 96; 103pp; French.  
 XX  
 CC The present sequence is exon 8 of human VG51 cDNA sequence, a glutamate/  
 CC gamma-aminobutyric acid (GABA) transporter. GABA and glutamate are  
 CC neurotransmitters. The transporter can be used to produce specific  
 CC antibodies to screen for binding agents. Modulators of the transporter  
 CC are useful for treating disorders associated with deregulated  
 CC glutamate/GABA transport, e.g. asthma, anxiety, epilepsy, hypertension  
 CC and other psychiatric and neurotic disorders, while determining levels of  
 CC the transporter and its coding sequence can be used for diagnosis of such  
 CC disorders. The full length coding sequence is given in ABR57930.  
 XX  
 SQ Sequence 167 BP; 40 A; 51 C; 37 G; 39 T; 0 other;  
 XX  
 QY Query Match 1.8%; Score 20; DB 23; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 6.7;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 324 GATGTCGACGATGACCATGC 343  
 DB 81 GATGTCGACGATGACCATGC 62  
 XX  
 RESULT 7  
 AAS90653/C  
 ID AAS90653 standard; cDNA; 739 BP.  
 XX  
 AC AAS90653;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #26457.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR  
 XX P-PSDB; ABG26466.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 1; SEQ ID No 26457; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue; as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 739 BP; 145 A; 231 C; 168 G; 193 T; 2 other;  
 XX  
 QY Query Match 1.8%; Score 20; DB 23; Length 739;  
 Best Local Similarity 100.0%; Pred. No. 6.2;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 324 GATGTCGACGATGACCATGC 343  
 DB 470 GATGTCGACGATGACCATGC 451  
 XX  
 RESULT 8  
 ABV23244/C  
 ID ABV23244 standard; cDNA; 947 BP.  
 XX  
 AC ABV23244;  
 XX  
 DT 16-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 23235.  
 XX  
 KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KM pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US05171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 XX  
 PR 16-MAR-2000; 2000US-189862P.  
 XX  
 PR 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI, 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1; Page 4190; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX  
 SQ Sequence 947 BP; 172 A; 307 C; 223 G; 241 T; 4 other;  
 XX  
 Query Match 1.8%; Score 20; DB 23; Length 947;  
 Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 324 GATGTGACGATGACCATGC 343  
 Db 470 GATGTGACGATGACCATGC 451  
 XX  
 RESULT 9  
 ABV29088/c  
 ID ABV29088 standard; cDNA; 947 BP.  
 XX  
 AC ABV29088;  
 XX  
 DT 16-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 29079.  
 XX  
 KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KM pharmacogenomic marker; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US05171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189622P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 09-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX

PI Schlegel R, Endege WO, Monahan JE;  
 XX  
 DR WPI, 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer -  
 XX  
 PS Claim 1; Page 6170-6171; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX  
 SQ Sequence 947 BP; 172 A; 307 C; 223 G; 241 T; 4 other;  
 XX  
 Query Match 1.8%; Score 20; DB 23; Length 947;  
 Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 324 GATGTGACGATGACCATGC 343  
 Db 470 GATGTGACGATGACCATGC 451  
 XX  
 RESULT 10  
 ABL57930/c  
 ID ABL57930 standard; cDNA; 1428 BP.  
 XX  
 AC ABL57930;  
 XX  
 DT 22-JUL-2002 (first entry)  
 XX  
 DE Human VG51 coding sequence.  
 XX  
 KM Antiasthmatic; anxiolytic; antiepileptic; antihypertensive; human;  
 KM psychotropic; glutamate transporter; transporter; GABA;  
 KM gamma-aminobutyric acid transporter; GABA transporter; neurotransmitter;  
 KM asthma; anxiety; epilepsy; hypertension; psychiatric disorder;  
 KM neurotic disorder; VG51; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1428  
 FT /\*tag= a  
 FT /partial  
 FT /product= "VG51"  
 FT /note= "No stop codon given"  
 XX  
 PN WO200071709-A1.  
 XX  
 PD 30-NOV-2000.  
 XX  
 PR 19-MAY-2000; 2000WO-FR01383.  
 XX  
 PR 21-MAY-1999; 99FR-0006525.  
 XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Giros B, Gasnier B, Sagne C, El Mestikawy S, Hamon M;

XX WP1; 2001-025160/03.  
 DR P-PSDB; ABB76941.  
 XX  
 PT New mammalian amino acid transporter, used e.g. to screen for  
 PT psychotropic agents, is high capacity but low affinity transporter of  
 PT gamma-aminobutyric acid -  
 PS Claim 5; Fig 13; 103pp; French.  
 XX  
 CC The present sequence is the coding sequence for human VGSL, a glutamate/  
 CC gamma-aminobutyric acid (GABA) transporter. GABA and glutamate are  
 CC neurotransmitters. The transporter can be used to produce specific  
 CC antibodies, to screen for binding agents. Modulators of the transporter  
 CC are useful for treating disorders associated with deregulated  
 CC glutamate/GABA transport, e.g. asthma, anxiety, epilepsy, hypertension  
 CC and other psychiatric and neurotic disorders, while determining levels of  
 CC the transporter and its coding sequence can be used for diagnosis of such  
 CC disorders.  
 XX  
 SQ Sequence 1428 BP; 289 A; 435 C; 337 G; 367 T; 0 other;  
 Query Match 1.8%; Score 20; DB 23; Length 1428;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 324 GATGGTGCATGACCATGC 343  
 Db 903 GATGGTGCATGACCATGC 884  
 RESULT 11  
 ABX70643/C  
 ID ABX70643 standard; cDNA; 1501 BP.  
 AC ABX70643;  
 XX  
 DT 04-MAR-2003 (first entry)  
 XX  
 DE Human cDNA encoding NOV7b.  
 XX  
 KW Human; ss; gene; NOVX; metabolic disorder; diabetes; cardiomyopathy;  
 KW obesity; infectious disease; anorexia; neurodegenerative disorder;  
 KW Alzheimer's disease; Parkinson's disease; immune disorder;  
 KW haematopoietic disorder; dyslipidaemia; metabolic disturbance;  
 KW metabolic syndrome X; wasting disorder; cancer; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200281518-A2.  
 XX  
 PD 17-OCT-2002.  
 XX  
 PF 21-FEB-2002; 2002WO-US05374.  
 XX  
 PR 21-FEB-2001; 2001US-270220P.  
 PR 21-FEB-2001; 2001US-270523P.  
 PR 23-FEB-2001; 2001US-270797P.  
 PR 23-FEB-2001; 2001US-270810P.  
 PR 08-MAR-2001; 2001US-274295P.  
 PR 16-MAR-2001; 2001US-276400P.  
 PR 16-MAR-2001; 2001US-276677P.  
 PR 26-MAR-2001; 2001US-278796P.  
 PR 04-APR-2001; 2001US-281521P.  
 PR 25-APR-2001; 2001US-286548P.  
 PR 13-AUG-2001; 2001US-311980P.  
 PR 10-SEP-2001; 2001US-318526P.  
 PR 17-SEP-2001; 2001US-322712P.  
 PR 18-OCT-2001; 2001US-330307P.  
 XX  
 PA (CUBRA-) CUBAGEN CORP.  
 XX  
 PI Pena CEA, Shinkets RA, Li L, Shenoy SG, Kekuda R, Spytek KA;

PI Vernet CM, Malyanar UL, Guo X, Gusev VY, Casman ST, Boldog FL;  
 PI Furtak K, Tchertnev VI, Patrajan M, Gangoli EA, Padigaru M;  
 PI Liu X, Baumgartner UC, Gerlach VJ, Spaderma SK, Zernusen BD;  
 XX  
 DR WP1; 2003-046859/04.  
 DR P-PSDB; ABUS2588.  
 XX  
 PT New isolated NOVX polypeptide useful for treating cardiomyopathy,  
 PT atherosclerosis, metabolic disorders, diabetes, obesity, infectious  
 PT disease, anorexia, neurodegenerative disorders, Alzheimer's disease and  
 PT cancer -  
 PS Claim 9; Page 57; 479pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide termed NOVX (NOV1,  
 CC 2a, 2b, 3a, 3b, 4a, 4b, 5, 6, 7a-c, 8a-e, 9a-b, 10, 11, 12a-c, 13, 14,  
 CC 15, 15a-d, 17a-b, 18, 19, 20a-b, 21-30) appearing as ABUS2578-ABUS2624),  
 CC a variant of NOVX, a mature form of NOVX, and a variant of the mature  
 CC form of NOVX. Also included are a nucleic acid molecule (NOVX NA)  
 CC encoding NOVX, or a fragment or complement of NOVX NA, a vector  
 CC comprising NOVX NA, a cell comprising the vector, an anti-NOVX antibody  
 CC (ab), determining the presence or amount of NOVX or NOVX NA in a sample,  
 CC and identifying an agent that binds or modulates the expression or  
 CC activity of NOVX. NOVX NA or ab is useful for treating or  
 CC preventing a NOVX-associated disorder in a subject, preferably human. Ab  
 CC is useful for determining the presence or amount of NOVX in a sample.  
 CC NOVX is useful for identifying an agent that binds to NOVX. NOVX  
 CC NA or ab is useful for treating metabolic disorders, diabetes,  
 CC cardiomyopathy, obesity, infectious disease, anorexia, neurodegenerative  
 CC disorders, Alzheimer's disease, Parkinson's disease, immune disorders,  
 CC haematopoietic disorders, and various dyslipidaemias, metabolic  
 CC disturbances associated with obesity, the metabolic syndrome X and  
 CC wasting disorders associated with chronic diseases, various cancers,  
 CC endocrine, connective tissue, blood, vascular, skin, renal, bone, brain,  
 CC muscle disorders, or bacterial, fungal, protozoal or viral infections.  
 CC NOVX, NOVX NA or ab is useful in screening assays, detection assays,  
 CC predictive medicine, and in methods of treatment. NOVX is useful as  
 CC immunogen, to screen for potential ant/agonist compounds, and as bait  
 CC protein in a two-hybrid or three-hybrid assay. NOVX NA is useful in gene  
 CC therapy, to express NOVX, to detect NOVX mRNA or a genetic lesion in a  
 CC NOVX gene, and to modulate NOVX activity. The cell is useful for  
 CC producing non-human transgenic animals. Ab is useful for isolating, and  
 CC purifying NOVX and to monitor protein levels in tissue as part of a  
 CC clinical testing procedure. The present sequence encodes a NOVX protein.  
 XX  
 SQ Sequence 1501 BP; 303 A; 459 C; 361 G; 378 T; 0 other;  
 Query Match 1.8%; Score 20; DB 25; Length 1501;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 324 GATGGTGCATGACCATGC 343  
 Db 964 GATGGTGCATGACCATGC 945  
 RESULT 12  
 ABX70642/C  
 ID ABX70642 standard; cDNA; 1513 BP.  
 AC ABX70642;  
 XX  
 DT 04-MAR-2003 (first entry)  
 XX  
 DE Human cDNA encoding NOV7a.  
 XX  
 KW Human; ss; gene; NOVX; metabolic disorder; diabetes; cardiomyopathy;  
 KW obesity; infectious disease; anorexia; neurodegenerative disorder;  
 KW Alzheimer's disease; Parkinson's disease; immune disorder;  
 KW haematopoietic disorder; dyslipidaemia; metabolic disturbance;  
 KW metabolic syndrome X; wasting disorder; cancer; gene therapy; SNP;  
 KW single nucleotide polymorphism.



CC systemic lupus erythematosus, rheumatoid arthritis, insulin-dependent  
CC diabetes mellitus); myeloid or lymphoid cell disorders (e.g. anaemia  
CC and thrombocytopenia); wounds, ulcers, burns; bone disorders (e.g.  
CC osteoporosis, osteoarthritis); mechanical and traumatic disorders (e.g.  
CC stroke, head trauma); lung or liver fibrosis; reperfusion injury in  
CC various tissues; bacterial, viral or fungal infections; allergic  
CC conditions such as allergic rhinitis, asthma, coagulation disorders  
CC (e.g. haemophilia); cancer and tumours; and inflammatory diseases (e.g.  
CC septic shock, Crohn's disease, anaphylaxis). The protein may be used to  
CC inhibit the growth, infection or function of infectious agents such as  
CC bacteria, fungi, viruses, or to effect bodily characteristics,  
CC biohythms or circadian cycles of rhythms. The protein may also  
CC have proliferation/differentiation, stem cell growth factor,  
CC haematopoiesis regulation, immune stimulating or suppressing,  
CC chemotactic/chemokinetic, hemostatic and thrombolytic, receptor/ligand,  
CC and antiinflammatory activities. The cDNA sequences of the invention are  
CC useful for expressing recombinant protein for analysis. The present  
CC sequence represents a novel human cDNA sequence of the invention,  
CC this sequence is an expressed sequence tag (EST) and was identified  
CC using subtractive hybridisation.

XX SQ Sequence 1597 BP; 327 A; 483 C; 383 G; 404 T; 0 other;

Query Match 1.8%; Score 20; DB 25; Length 1597;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTGACGATGACCATGC 343  
DB 1005 GATGTGACGATGACCATGC 986

## RESULT 14

AA590655  
ID AA590655 standard; cDNA; 1771 BP.

XX AC AA590655;

XX DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #26459.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HXSSE-) HXSSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX DR P-PSDB; ABG26468.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity -

XX PS Claim 1; SEQ ID No 26459; 103bp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AA564197-AA594564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 1771 BP; 422 A; 444 C; 529 G; 376 T; 0 other;

Query Match 1.8%; Score 20; DB 23; Length 1771;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTGACGATGACCATGC 343  
DB 1347 GATGTGACGATGACCATGC 1366

## RESULT 15

AB280236/C  
ID AB280236 standard; cDNA; 2057 BP.

XX AC AB280236;

XX DT 02-JUN-2003 (first entry)

DE Human transdormin 3 encoding cDNA SEQ ID NO:5.

XX KW Neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy;

XX KW central nervous system disorder; CNS disorder; multiple sclerosis;

XX KW nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; transdormin; human; transdormin 3; chromosome 5; gene; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 152..1582

XX FT /tag=a

XX FT /product="transdormin 3"

XX FT /transl\_except="(pos:1217..1219,aa:11e)

XX PN WO2003016502-A2.

XX PD 27-FEB-2003.

PF 21-AUG-2002; 2002WO-US26637.

XX PR 21-AUG-2001; 2001US-313907P.

XX PR 21-AUG-2002; 2002US-0225810.

XX PA (MCLA-) MCLAUGHLIN RES INST.

XX PI Birmingham JR;

XX DR WPI; 2003-278567/27.

XX DR P-PSDB; ABP96441.

XX PT New nucleic acid sequence encoding transdormin, e.g. mouse transdormin 1,

XX PT mouse transdormin 2, human transdormin 1, human transdormin 2, human transdormin 3 or rat transdormin 1, useful for treating CNS, e.g. stroke, multiple sclerosis, trauma, neuropathic pain -

XX  
PS Claim 1; Fig 21; 177pp; English.  
XX  
CC The present invention describes an isolated nucleic acid sequence  
CC comprising a cDNA sequence encoding mouse tramdorin (trandl 2, mouse  
CC trand 3, human trand 1, human trand 2, human trand 3 or rat trand 1, or  
CC the genomic sequence of mouse trand 1 or mouse trand 3. Mouse trand 1 is  
CC located to chromosome 11, whereas human trand 1 is located to chromosome  
CC 5q31-33. The trand sequences have neuroprotective, nootropic, analgesic  
CC and cerebroprotective activities, and can be used in gene therapy. The  
CC nucleic acid sequences are useful for diagnosing and treating central  
CC nervous system (CNS) disorders such as multiple sclerosis, nerve injury,  
CC neuropathic pain, stroke or trauma, and non-CNS disorders. The present  
CC sequence encodes human trand 3, which is given in the exemplification of  
CC the present invention.  
XX

SQ Sequence 2057 BP; 411 A; 638 C; 503 G; 505 T; 0 other;

Query Match 1.8%; Score 20; DB 25; Length 2057;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTCAGCATGACCATGC 343  
DB 1054 GATGTCAGCATGACCATGC 1035

Search completed: November 8, 2003, 21:40:10  
Job time : 283 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 21:30:02 ; Search time 70 Seconds  
(without alignments)

7188.240 Million cell updates/sec

Title: US-10-007-527a-1

Perfect score: 1140

Sequence: 1 atgaccagcgaagtcgtcga.....taatatcggtcgaatrga 1140

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/ECTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	1.7	4580	2	US-08-674-351-1
2	18	1.6	4975	2	US-08-249-687C-1
3	18	1.6	4989	2	US-08-666-392A-3
4	18	1.6	4989	2	US-08-625-819-1
5	18	1.6	4989	3	US-08-755-558-4
6	18	1.6	4989	3	US-08-746-559A-1
7	18	1.6	4989	3	US-08-880-313A-9
8	18	1.6	4989	3	US-09-199-926-3
9	18	1.6	4989	4	US-08-864-641B-17
10	18	1.6	4989	4	US-09-389-855A-9
11	18	1.6	4989	4	US-09-668-822-9
12	18	1.6	4993	3	US-08-746-559A-3
13	18	1.6	4403765	3	US-09-103-840A-2
14	18	1.6	4411529	3	US-09-103-840A-1
15	17	1.5	5433	4	US-09-252-991A-12481
16	17	1.5	705	4	US-09-252-991A-12481
17	17	1.5	1653	3	US-09-230-944-19
18	17	1.5	1653	3	US-08-881-784-8
19	17	1.5	1653	3	US-09-292-768-3
20	17	1.5	1653	3	US-09-292-768-67
21	17	1.5	1653	3	US-09-292-768-69
22	17	1.5	1687	1	US-08-279-700-17
23	17	1.5	1687	1	US-08-279-700-19
24	17	1.5	1762	3	US-08-881-784-5
25	17	1.5	1762	3	US-09-292-768-1
26	17	1.5	1762	3	US-09-292-768-63
27	17	1.5	1762	3	US-09-292-768-65

28	17	1.5	1762	3	US-09-172-339-5	Sequence 5, Appli
29	17	1.5	2352	4	US-08-997-251-3	Sequence 3, Appli
30	17	1.5	48908	4	US-09-453-702B-137	Sequence 137, App
31	17	1.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
32	17	1.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
33	16	1.4	139	4	US-09-141-027-11	Sequence 11, Appli
34	16	1.4	139	4	US-09-617-804-11	Sequence 11, Appli
35	16	1.4	219	2	US-08-414-657D-11	Sequence 11, Appli
36	16	1.4	219	2	US-08-414-657D-12	Sequence 11, Appli
37	16	1.4	276	4	US-09-673-809-23	Sequence 73, Appli
38	16	1.4	276	4	US-09-673-809-73	Sequence 73, Appli
39	16	1.4	340	2	US-08-997-080-137	Sequence 137, App
40	16	1.4	340	2	US-08-997-362-137	Sequence 137, App
41	16	1.4	340	3	US-09-095-855-137	Sequence 137, App
42	16	1.4	340	4	US-09-324-542-137	Sequence 137, App
43	16	1.4	340	4	US-09-205-426-137	Sequence 137, App
44	16	1.4	468	4	US-09-370-838-133	Sequence 133, App
45	16	1.4	504	4	US-09-134-001C-2229	Sequence 2229, Ap

## ALIGNMENTS

RESULT 1  
US-08-674-351-1/C  
Sequence 1, Application US/08674351  
Patent No. 5831013  
GENERAL INFORMATION:  
APPLICANT: Bruenn, Jeremy A.  
TITLE OF INVENTION: CAPSID POLYPEPTIDES AND USE TO INHIBIT  
TITLE OF INVENTION: VIRAL PACKAGING  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/674,351  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Timian, Susan J.  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 19226/740  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1636  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4580 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: CDNA  
US-08-674-351-1

Query Match 1.7% Score 19; DB 2; Length 4580;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 541 GTGAGTGAGACATCTCG 559  
|||||  
DB 3310 GTGAGTGAGACATCTCG 3292

RESULT 2  
US-08-249-687C-1/c  
; Sequence 1, Application US/08249687C  
; Patent No. 5942412  
; GENERAL INFORMATION:  
; APPLICANT: PRAGER, DIANE  
; APPLICANT: MELMED, SHLOMO  
; TITLE OF INVENTION: POLYNUCLEIC ACID ENCODING  
; TITLE OF INVENTION: VARIANT INSULIN-LIKE GROWTH FACTOR I RECEPTOR BETA  
; TITLE OF INVENTION: SUBUNIT & RECEPTOR  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder & Poplawski  
; STREET: 444 South Flower St., 19th Floor  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/249,687C  
; FILING DATE: 26-MAY-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/044,540  
; FILING DATE: 06-APR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Basile, Lena  
; REGISTRATION NUMBER: P-44,026  
; REFERENCE/DOCKET NUMBER: P07 32249  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 213-622-7700  
; TELEFAX: 213-489-4210  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4975 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: both  
; MOLECULE TYPE: cDNA  
US-08-249-687C-1  
  
Query Match 1.6%; Score 18; DB 2; Length 4975;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1039 CTCGTCGAAGCTGGCGCG 1056  
DB 4057 CTCGTCGAAGCTGGCGCG 4040

RESULT 3  
US-08-666-392A-3/c  
; Sequence 3, Application US/08666392A  
; Patent No. 5928040  
; GENERAL INFORMATION:  
; APPLICANT: (countries other than U.S.): Royal Children's Hospital  
; APPLICANT: Research Foundation  
; APPLICANT: (U.S. only): George A. WERTHER and  
; APPLICANT: Christopher J. WRIGHT  
; TITLE OF INVENTION: A METHOD FOR THE PROPHYLAXIS AND/OR  
; TITLE OF INVENTION: TREATMENT OF PROLIFERATIVE AND/OR  
; TITLE OF INVENTION: INFLAMMATORY SKIN DISORDERS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza

CITY: New York  
STATE: NY  
COUNTRY: U.S.A.  
ZIP: 10112-0228  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/666,392A  
FILING DATE: 20-AUG-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU95/00410  
FILING DATE: 06-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: MacLeod, Janet M.  
REGISTRATION NUMBER: 35,263  
REFERENCE/DOCKET NUMBER: A30626-PCT-USA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-408-2500  
TELEFAX: 212-765-2519  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4989 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
PUBLICATION INFORMATION:  
AUTHORS: Ulrich, A., et al.  
TITLE: Insulin-like growth factor I receptor...  
JOURNAL: EMBO J.  
VOLUME: 5  
ISSUE: 1986  
PAGES: 2503-2512  
DATE:  
US-08-666-392A-3  
  
Query Match 1.6%; Score 18; DB 2; Length 4989;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1039 CTCGTCGAAGCTGGCGCG 1056  
DB 4071 CTCGTCGAAGCTGGCGCG 4054

RESULT 4  
US-08-625-819-1/c  
; Sequence 1, Application US/08625819  
; Patent No. 5958872  
; GENERAL INFORMATION:  
; APPLICANT: O'CONNOR, Rosemary; and  
; APPLICANT: BASERGA, Renato L.  
; TITLE OF INVENTION: ACTIVE SURVIVAL DOMAINS OF IGF-IR  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE and DORR LLP  
; STREET: 1455 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

```

? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/625,819
? FILING DATE: 01-APR-1996
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: WIXON, Henry N.
? REGISTRATION NUMBER: 32,073
? REFERENCE/DOCKET NUMBER: 104322.162
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 942-8459
? TELEFAX: (202) 942-8484
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4989 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 46..4149
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 136..4149
? US-08-625-819-1

Query Match 1.6%; Score 18; DB 2; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGCG 1056
DB 4071 CTCGTGAAGCTGGCGCG 4054

RESULT 5
US-08-755-558-4/C
? Sequence 4, Application US/08755558
? Patent No. 6071891
? GENERAL INFORMATION:
? APPLICANT: Low, Walter
? APPLICANT: Wallentriedman, Margaret
? APPLICANT: Chiang, Lan
? TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTOR 1 RECEPTOR (IGF-1R) ANTISEN
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
? STREET: 3100 No. 6071891west Center, 90 S. 7th Street
? CITY: Minneapolis
? STATE: MN
? COUNTRY: U.S.A.
? ZIP: 55402
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq Version 1.5
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/755,558
? FILING DATE: 20-NOV-1996
? CLASSIFICATION: 424
? PRIOR APPLICATION NUMBER:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Kettleberger, Denise M
? REGISTRATION NUMBER: 33,924
? REFERENCE/DOCKET NUMBER: 600.337US01
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 612/332-5300
? TELEFAX: 612/332-9081
```

```

? TELEX:
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4989 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: Genomic DNA
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? FRAGMENT TYPE:
? ORIGINAL SOURCE:
? FEATURE:
? NAME/KEY: Coding Sequence
? LOCATION: 46..4149
? OTHER INFORMATION:
? US-08-755-558-4

Query Match 1.6%; Score 18; DB 3; Length 4989;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGCG 1056
DB 4071 CTCGTGAAGCTGGCGCG 4054

RESULT 6
US-08-746-559A-1/C
? Sequence 1, Application US/08746559A
? Patent No. 6084085
? GENERAL INFORMATION:
? APPLICANT: Renato Baserga
? APPLICANT: Mariana Resnicoff
? APPLICANT: Consuelo D'Ambrosio
? APPLICANT: Andre Ferber
? TITLE OF INVENTION: Method of Inducing Resistance to Tumor Growth
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6084085r1s LLP
? STREET: One Liberty Place - 46th Floor
? CITY: Philadelphia
? STATE: PA
? COUNTRY: USA
? ZIP: 19103
? COMPUTER READABLE FORM:
? MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: PC-DOS
? SOFTWARE: WORDPERFECT 6.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/746,559A
? FILING DATE: 13-NOV-1996
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 60/006,699
? FILING DATE: 14-NOV-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Paul K. Legaard
? REGISTRATION NUMBER: 38,534
? REFERENCE/DOCKET NUMBER: TUT-2063
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (215) 568-3100
? TELEFAX: (215) 568-3439
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4989 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? US-08-746-559A-1

Query Match 1.6%; Score 18; DB 3; Length 4989;
```

Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGG 1056  
|||||  
Db 4071 CTCGTGAAGCTGGCGG 4054

## RESULT 7

US-08-880-313A-9/c  
; Sequence 9, Application US/08880313A  
; Patent No. 6274562  
; GENERAL INFORMATION:  
; APPLICANT: Baserga, Renato  
; APPLICANT: Sell, Christian  
; APPLICANT: Rubin, Raphael  
; TITLE OF INVENTION: Method of Inhibiting the Proliferation and  
; TITLE OF INVENTION: Causing the Differentiation of Cells with IGF-1 Receptor  
; TITLE OF INVENTION: Antisense Oligonucleotides  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Woodcock Washburn Kurtz Maclewiecz & No. 6274562xis LLP  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: U.S.A.  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 8.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/880,313A  
; FILING DATE: June 20, 1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/479,173  
; FILING DATE: June 6, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul K. Leggaard  
; REGISTRATION NUMBER: 38,534  
; REFERENCE/DOCKET NUMBER: TUD-2383  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-568-3100  
; TELEFAX: 215-568-3439  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4989 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-880-313A-9

Query Match 1.6%; Score 18; DB 3; Length 4989;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGG 1056  
|||||  
Db 4071 CTCGTGAAGCTGGCGG 4054

RESULT 8  
US-09-199-926-3/c  
; Sequence 3, Application US/09199926  
; Patent No. 6284741  
; GENERAL INFORMATION:  
; APPLICANT: (countries other than U.S.): Royal Children's Hospital  
; APPLICANT: Research Foundation  
; APPLICANT: (U.S. only): George A. WERTHER and  
; APPLICANT: Christopher J. WRIGHT  
; TITLE OF INVENTION: A METHOD FOR THE PROPHYLAXIS AND/OR

TITLE OF INVENTION: TREATMENT OF PROLIFERATIVE AND/OR  
; TITLE OF INVENTION: INFLAMMATORY SKIN DISORDERS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: NY  
; COUNTRY: U.S.A.  
; ZIP: 10112-0228  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/199,926  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/666,392  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Macleod, Janet M.  
; REGISTRATION NUMBER: 35,263  
; REFERENCE/DOCKET NUMBER: A30626-PCT-USA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-408-2500  
; TELEFAX: 212-765-2519  
; TELEX:

## INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
; LENGTH: 4989 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; PUBLICATION INFORMATION:  
; AUTHORS: Ulrich, A., et al.  
; TITLE: Insulin-like growth factor I receptor...  
; JOURNAL: EMBO J.  
; VOLUME: 5  
; ISSUE: 1986  
; PAGES: 2503-2512  
; DATE:  
; US-09-199-926-3

Query Match 1.6%; Score 18; DB 3; Length 4989;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTGAAGCTGGCGG 1056  
|||||  
Db 4071 CTCGTGAAGCTGGCGG 4054

## RESULT 9

US-08-864-641B-17/c  
; Sequence 17, Application US/08864641B  
; Patent No. 6312684  
; GENERAL INFORMATION:  
; APPLICANT: Baserga, Renato  
; APPLICANT: Abraham, David  
; APPLICANT: Resnicoff, Mariana  
; TITLE OF INVENTION: Method of Inducing Resistance To Tumor Growth  
; FILE REFERENCE: TUD2137  
; CURRENT APPLICATION NUMBER: US/08/864,641B  
; CURRENT FILING DATE: 1997-05-29  
; PRIOR APPLICATION NUMBER: 08/340,732

PRIOR FILING DATE: 1994-11-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 17  
LENGTH: 4989  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: No. 6312684el Sequence  
NAME/KEY: CDS  
LOCATION: (46)..(4146)  
US-08-864-641B-17

Query Match 1.6%; Score 18; DB 4; Length 4989;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTCGAAGCTGGCGCG 1056  
DB 4071 CTCGTCGAAGCTGGCGCG 4054

RESULT 10  
US-09-389-855A-9/c  
Sequence 9, Application US/09389855A  
Patent No. 6331526  
GENERAL INFORMATION:  
APPLICANT: Baserga, Renato  
APPLICANT: Sell, Christian  
APPLICANT: Rubin, Raphael  
TITLE OF INVENTION: Method Of Inhibiting The Proliferation And Causing The  
TITLE OF INVENTION: Differentiation Of Cells With IGF-1 Receptor Antisense  
FILE REFERENCE: TJU-2388  
CURRENT APPLICATION NUMBER: US/09/389,855A  
CURRENT FILING DATE: 1999-09-02  
PRIOR APPLICATION NUMBER: US 08/880,313  
PRIOR FILING DATE: 1997-06-20  
PRIOR APPLICATION NUMBER: US 08/479,173  
PRIOR FILING DATE: 1995-06-06  
PRIOR APPLICATION NUMBER: US 08/158,761  
PRIOR FILING DATE: 1993-11-30  
PRIOR APPLICATION NUMBER: US 08/037,257  
PRIOR FILING DATE: 1993-03-26  
NUMBER OF SEQ ID NOS: 9  
SEQ ID NO 9  
LENGTH: 4989  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Other Information: human IGF-1R sequence  
US-09-389-855A-9

Query Match 1.6%; Score 18; DB 4; Length 4989;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTCGAAGCTGGCGCG 1056  
DB 4071 CTCGTCGAAGCTGGCGCG 4054

RESULT 11  
US-09-668-822-9/c  
Sequence 9, Application US/09668822  
Patent No. 6340674  
GENERAL INFORMATION:  
APPLICANT: Baserga, Renato  
APPLICANT: Sell, Christian  
APPLICANT: Rubin, Raphael  
TITLE OF INVENTION: Method of Inhibiting the Proliferation and Causing the Differ  
NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6340674ris LLP  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: U.S.A.  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/668,822  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/880,313  
FILING DATE: June 20, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul K. Legaard  
REGISTRATION NUMBER: 38,534  
REFERENCE/DOCKET NUMBER: TJU-2443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-568-3100  
TELEFAX: 215-568-3439  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4989 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-668-822-9

Query Match 1.6%; Score 18; DB 4; Length 4989;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1039 CTCGTCGAAGCTGGCGCG 1056  
DB 4071 CTCGTCGAAGCTGGCGCG 4054

RESULT 12  
US-08-746-559A-3/c  
Sequence 3, Application US/08746559A  
Patent No. 6084085  
GENERAL INFORMATION:  
APPLICANT: Renato Baserga  
APPLICANT: Mariana Resnicoff  
APPLICANT: Consuelo D'Ambrosio  
APPLICANT: Andre Ferber  
TITLE OF INVENTION: Method of Inducing Resistance to Tumor Growth  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6084085ris LLP  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/746,559A  
FILING DATE: 13-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/006,699  
FILING DATE: 14-NOV-1995

```
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Paul K. Legaard
/ REGISTRATION NUMBER: 38,534
/ REFERENCE/DOCKET NUMBER: TJU-2063
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (215) 568-3100
/ TELEFAX: (215) 568-3439
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 4993 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ US-08-746-559A-3

Query Match      1.6%; Score 18; DB 3; Length 4993;
Best Local Similarity 100.0%; Pred.No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1039 CTCGTCGAAGCTGGCGC 1056
         |||||
Db      4075 CTCGTGGAAGCTGGCGCG 4058

RESULT 13
US-09-103-840A-2
/ Sequence 2, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      1.6%; Score 18; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred.No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      759 ACATGGCAACCGTGACCC 776
         |||||
Db      2656723 ACATGGCAACCGTGACCC 2656740

RESULT 14
US-09-103-840A-1
/ Sequence 1, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
```

```
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 4411529
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      1.6%; Score 18; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred.No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      759 ACATGGCAACCGTGACCC 776
         |||||
Db      2659427 ACATGGCAACCGTGACCC 2659444

RESULT 15
US-09-252-991A-12481
/ Sequence 12481, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 12481
/ LENGTH: 543
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12481

Query Match      1.5%; Score 17; DB 4; Length 543;
Best Local Similarity 100.0%; Pred.No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      132 CGCCTGTGGCGCGCCGA 148
         |||||
Db      422 CGCCTGTGGCGCGCCGA 438
```

Search completed: November 8, 2003, 23:04:09  
Job time : 87 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 21:35:22 ; Search time 304 Seconds  
(without alignments)  
11966.092 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1140  
1 atgaccagcgtaagtcgta.....taatatcggttcgcaagta 1140

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2141354 seqs, 1595478879 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA.\*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	1140	100.0	1140	US-10-007-527A-1	Sequence 1, Appli
2	1140	100.0	1140	US-10-007-452-1	Sequence 1, Appli
3	1140	100.0	6334	US-10-007-527A-5	Sequence 5, Appli
4	1140	100.0	6334	US-10-007-452-5	Sequence 5, Appli
5	1140	100.0	9652	US-10-007-527A-7	Sequence 7, Appli
6	1140	100.0	9652	US-10-007-452-7	Sequence 7, Appli
7	1140	100.0	11241	US-10-007-527A-6	Sequence 6, Appli
8	1140	100.0	11241	US-10-007-452-6	Sequence 6, Appli
9	20	1.8	20	US-10-007-527A-19	Sequence 19, Appli
10	20	1.8	20	US-10-007-452-19	Sequence 19, Appli
11	20	1.8	454	US-10-027-632-94798	Sequence 94798, A
12	20	1.8	454	US-10-027-632-94798	Sequence 94798, A
13	20	1.8	2055	US-10-225-810-5	Sequence 5, Appli
14	20	1.8	2093	US-09-805-456-1	Sequence 1, Appli
15	20	1.8	46649	US-09-805-456-3	Sequence 3, Appli
16	20	1.8	77992	US-10-225-810-11	Sequence 11, Appli

17	20	1.8	250000	US-10-225-810-26	Sequence 26, Appli
18	19	1.7	416	US-09-783-590-7376	Sequence 7376, Ap
19	19	1.7	447	US-09-918-995-9021	Sequence 9021, Ap
20	19	1.7	476	US-09-878-574-4387	Sequence 4387, Ap
21	19	1.7	1333	US-10-106-698-1797	Sequence 1797, Ap
22	19	1.7	3810	US-09-880-107-3709	Sequence 3709, Ap
23	18	1.6	20	US-10-007-527A-20	Sequence 20, Appli
24	18	1.6	20	US-10-007-452-20	Sequence 20, Appli
25	18	1.6	282	US-09-878-574-7769	Sequence 7769, Ap
26	18	1.6	399	US-09-764-869-1382	Sequence 1382, Ap
27	18	1.6	399	US-10-091-504-1382	Sequence 1382, Ap
28	18	1.6	440	US-09-764-869-64	Sequence 64, Appli
29	18	1.6	440	US-10-091-504-64	Sequence 64, Appli
30	18	1.6	1095	US-09-815-242-6193	Sequence 6193, Ap
31	18	1.6	1214	US-10-378-393-4	Sequence 4, Appli
32	18	1.6	1995	US-10-378-393-6	Sequence 6, Appli
33	18	1.6	4989	US-09-870-759-119	Sequence 119, App
34	18	1.6	4989	US-09-791-196-3	Sequence 3, Appli
35	18	1.6	4989	US-10-007-926A-130	Sequence 130, App
36	18	1.6	4989	US-09-751-708A-119	Sequence 119, App
37	18	1.6	4989	US-10-177-293-225	Sequence 225, App
38	18	1.6	8858	US-10-378-393-1	Sequence 1, Appli
39	17	1.5	113	US-09-969-373-620	Sequence 620, App
40	17	1.5	253	US-09-923-876-1088	Sequence 1088, Ap
41	17	1.5	343	US-09-734-563-109	Sequence 109, App
42	17	1.5	384	US-09-878-574-1881	Sequence 1881, Ap
43	17	1.5	509	US-10-029-386-11554	Sequence 11554, A
44	17	1.5	550	US-09-991-936-1453	Sequence 1453, Ap
45	17	1.5	561	US-10-027-632-290199	Sequence 290199,

## ALIGNMENTS

```
RESULT 1
US-10-007-527A-1
; Sequence 1, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kowitnicka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; CURRENT FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 1
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Rhodococcus AN12
US-10-007-527A-1
```

Query Match 100.0%; Score 1140; DB 14; length 1140;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACCAGCGTAAGTGTGAACCTTTCCGGGAAAGACCGGCTCCCTCGTGTG 60
1 ATGACCAGCGTAAGTGTGAACCTTTCCGGGAAAGACCGGCTCCCTCGTGTG 60
Db 1 ATGACCAGCGTAAGTGTGAACCTTTCCGGGAAAGACCGGCTCCCTCGTGTG 60
QY 61 TCCGATTAACCGCGGCAATCCGGGACGAACTTCAAGTAATCAACAGTCA 120
61 TCCGATTAACCGCGGCAATCCGGGACGAACTTCAAGTAATCAACAGTCA 120
Db 61 TCCGATTAACCGCGGCAATCCGGGACGAACTTCAAGTAATCAACAGTCA 120
QY 121 GAAACATTTAAGCCTGTGCGCGCGCGATTTTGGCGTGAACGATGTCAAC 180
121 GAAACATTTAAGCCTGTGCGCGCGCGATTTTGGCGTGAACGATGTGTCAAC 180
Db 121 GAAACATTTAAGCCTGTGCGCGCGCGATTTTGGCGTGAACGATGTGTCAAC 180
```



```

RESULT 2
US-10-007-452-1
; Sequence 1, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong

```

ORGANISM: Rhodococcus AN12  
US-10-007-452-1

Best Local Similarity	100.0%	Pred. No.	0					
Matches 1140	Conservative	0	Mismatches	0	Indels	0	Gaps	0

QY	1	ATGACGAGCGTAAGTGTGGAACACCTTTCGCGCAAAAGCGGGCTCCGCTCTCGTGTGCG	60
Db	1	ATGACGAGCGTAAGTGTGGAACACCTTTCGCGCAAAAGCGGGCTCCGCTCTCGTGTGCG	60
QY	61	TCCGATAGAGCGCGGCATCCCGCACGAACGTGGACCCAACTTCAACAAATACACGTCGA	120
Db	61	TCCGATAGAGCGCGGCATCCCGCACGAACGTGGACCCAACTTCAACAAATACACGTCGA	120
QY	121	GAAACATTTAAACGCTGTGGCCGGCCGATTTCTGGGCGTGAACGGGTGGACCATTTGTGCAC	180
Db	121	GAAACATTTAAACGCTGTGGCCGGCCGATTTCTGGGCGTGAACGGGTGGACCATTTGTGCAC	180
QY	181	GGTCCGAAAAGGTTCTGGATTCGAGAGGCGCTTGTTCTCGCGAAAAGGCGTGATCTGCCCC	240
Db	181	GGTCCGAAAAGGTTCTGGATTCGAGAGGCGCTTGTTCTCGCGAAAAGGCGTGATCTGCCCC	240
QY	241	TGCTGTGCGGAAAAGTCGGTGCACATCGTGCAGACGAAATTTCTCAAGTTGTGTCTCAT	300
Db	241	TGCTGTGCGGAAAAGTCGGTGCACATCGTGCAGACGAAATTTCTCAAGTTGTGTCTCAT	300
QY	301	CAACTCGGGACTGGAATCTGTTGGCATGTGGACATGACCATGCGGCATACAGCTGTGAG	360
Db	301	CAACTCGGGACTGGAATCTGTTGGCATGTGGACATGACCATGCGGCATACAGCTGTGAG	360
QY	361	CGGCTCCACGACCTATGGACTGACTTTGGCAGCGTGGAAAGCGTGCACCAACGAGTGT	420
Db	361	CGGCTCCACGACCTATGGACTGACTTTGGCAGCGTGGAAAGCGTGCACCAACGAGTGT	420
QY	421	CGTTGGCGTACGGAACGTGAATGTACGGCTGCACGGAATGCGTGGCGCTGTGGAAATC	480
Db	421	CGTTGGCGTACGGAACGTGAATGTACGGCTGCACGGAATGCGTGGCGCTGTGGAAATC	480
QY	481	ACTCACGGAATAAAACGGCTGGCACGTCACGTTCAACGCGCTACATGTTCAGTGTGAC	540
Db	481	ACTCACGGAATAAAACGGCTGGCACGTCACGTTCAACGCGCTACATGTTCAGTGTGAC	540
QY	541	GTTGATGGGAACATCTCTCGAATCCTTTCGCGATGGATGGATGGTTCGATCCGCTCCAA	600
Db	541	GTTGATGGGAACATCTCTCGAATCCTTTCGCGATGGATGGATGGTTCGATCCGCTCCAA	600
QY	601	CTCGATCTCGGGAATTTGCTGCGCCACTAGATTAATCGGAGGTCTCGATATGCAAAAG	660
Db	601	CTCGATCTCGGGAATTTGCTGCGCCACTAGATTAATCGGAGGTCTCGATATGCAAAAG	660
QY	661	ATCGCGGTGAAGCTGATCAAGTTCTCGCTGCGTATCTGACGAAAATTTGCATCTGCGTT	720
Db	661	ATCGCGGTGAAGCTGATCAAGTTCTCGCTGCGTATCTGACGAAAATTTGCATCTGCGTT	720
QY	721	GGTATGAGGTTGGTGAATGGTGGACGGAATAAAAGTGTGCACATGGCAACCTGTGG	780
Db	721	GGTATGAGGTTGGTGAATGGTGGACGGAATAAAAGTGTGCACATGGCAACCTGTGG	780
QY	781	GAAATCGTGTGAAGTGGAGGGCGGGGATCCAAAGGTTGGAACGTGGCGGAATTT	840
Db	781	GAAATCGTGTGAAGTGGAGGGCGGGGATCCAAAGGTTGGAACGTGGCGGAATTT	840

Db 781 GAATCGCTGTTGATGACGTGGGCGGGATCCACAAGCTTGGAACTGTGGCGAGAAATT 840  
Qy 841 GAGTTGGTTCGATGAGAGCTCGGGCAATCGCGTGGTCCCGTGAATTCGTCGCGAGCT 900  
Db 841 GAGTTGGTTCGATGAGAGCTCGGGCAATCGCGTGGTCCCGTGAATTCGTCGCGAGCT 900  
Qy 901 GGTCTTGGGGCGAAGCTAAACAGATGCTCAGATCGTTGAGCAGAGAAATCTGCCCGGCTC 960  
Db 901 GGTCTTGGGGCGAAGCTAAACAGATGCTCAGATCGTTGAGCAGAGAAATCTGCCCGGCTC 960  
Qy 961 ATGTTGGGATTCATTCGCGCGGCGATCGTGAATGATTCGCACTTGGCGCGCTTAAGTC 1020  
Db 961 ATGTTGGGATTCATTCGCGCGGCGATCGTGAATGATTCGCACTTGGCGCGCTTAAGTC 1020  
Qy 1021 TTCGCGAGATCCTCGAGCTCGTCGAAAGCTGCGCGCACTTGGGAAATCTTCGTATCAC 1080  
Db 1021 TTCGCGAGATCCTCGAGCTCGTCGAAAGCTGCGCGCACTTGGGAAATCTTCGTATCAC 1080  
Qy 1081 TTGCATTTATCGATTGCCCGCAGCGGATGTGCGGCCCGCATATATCGGTTGCAAGTGA 1140  
Db 1081 TTGCATTTATCGATTGCCCGCAGCGGATGTGCGGCCCGCATATATCGGTTGCAAGTGA 1140

RESULT 3  
US-10-007-527a-5/c  
; Sequence 5, Application US/1007527A  
; Publication No. US2003004807A1  
; GENERAL INFORMATION:

; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: Kostichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: CI1709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,527A  
; PRIOR FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: 60/254,868  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 5  
; LENGTH: 6334  
; TYPE: DNA  
; ORGANISM: Rhodococcus AN12  
US-10-007-527a-5

Query Match 100.0%; Score 1140; DB 14; Length 6334;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGACGAGGTAGTGTCTGAACACCTTCCGGCAAGACCGGCTCCCGTCCGTGTCG 60  
Db 3051 ATGACGAGGTAGTGTCTGAACACCTTCCGGCAAGACCGGCTCCCGTCCGTGTCG 2992  
Qy 61 TTCGATTAAGCGCGCATCCGCGACGAATGCGACCCAACTTCAACAAATCAACAGTCA 120  
Db 2991 TTCGATTAAGCGCGCATCCGCGACGAATGCGACCCAACTTCAACAAATCAACAGTCA 2932  
Qy 121 GAAACATTTAACCGCTGTGGCGCGCGCGATTTCTGGCGTGAACGCTGTACATTGTCAAC 180  
Db 2931 GAAACATTTAACCGCTGTGGCGCGCGCGATTTCTGGCGTGAACGCTGTGTCAAC 2872  
Qy 181 GGTCCGAAAAGTCTGTGATTCGGAAGCTTCCTCCGCGGAAAGGCTGATCTGCCCC 240  
Db 2871 GGTCCGAAAAGTCTGTGATTCGGAAGCTTCCTCCGCGGAAAGGCTGATCTGCCCC 2812  
Qy 241 TGTCTGTGGGAAAAGTGTGATTCGTCGATCGTCGAGACGAAATTTCTCAAGTTGTGTCTAT 300  
Db 2811 TGTCTGTGGGAAAAGTGTGATTCGTCGATCGTCGAGACGAAATTTCTCAAGTTGTGTCTAT 2752  
Qy 301 CAACCTCGGAGCTGTGATTCGTCGATGTGATCAATGACATGCGCCATACAGCTGTCTAG 360  
Db 2751 CAACCTCGGAGCTGTGATTCGTCGATGTGATCAATGACATGCGCCATACAGCTGTCTAG 2692

Qy 361 CGGCTCCACGACCTATGATCTGACCTTTCCGACGCTTGAAAGCTCGGACCTAACGCTCT 420  
Db 2691 CGGCTCCACGACCTATGATCTGACCTTTCCGACGCTTGAAAGCTCGGACCTAACGCTCT 2632  
Qy 421 CGTTGGGCTAACGAAAGTGAATGTAACGCGTGCACGGAATCGTGGCGCTGTTGAAATC 480  
Db 2631 CGTTGGGCTAACGAAAGTGAATGTAACGCGTGCACGGAATCGTGGCGCTGTTGAAATC 2572  
Qy 481 ACTCACGGAAGAAAACGCGTGCACGTCACAGTTACGCGCTACTCATGTTGATGATGAC 540  
Db 2571 ACTCACGGAAGAAAACGCGTGCACGTCACAGTTACGCGCTACTCATGTTGATGATGAC 2512  
Qy 541 GTGATGAGAACATCTCGTAATCTTCTCGATGATGATGATGATGATGATGATGATGAT 600  
Db 2511 GTGATGAGAACATCTCGTAATCTTCTCGATGATGATGATGATGATGATGATGATGAT 2452  
Qy 601 CTGCTATCTCTGGATTTGCTGCGCCACATACGTAATGGGCTGGTCTGATGATGATGAT 660  
Db 2451 CTGCTATCTCTGGATTTGCTGCGCCACATACGTAATGGGCTGGTCTGATGATGATGAT 2392  
Qy 661 ATCGCGGTGAAGCTGATCAAGTTCTCGCTCGTATCTGACGAAATTCATCTGGCGTT 720  
Db 2391 ATCGCGGTGAAGCTGATCAAGTTCTCGCTCGTATCTGACGAAATTCATCTGGCGTT 2332  
Qy 721 GGTATGAGAGTTGTGATGTCGCGACGAGAAAAGTGTGACATGTCGACACCTGTCG 780  
Db 2331 GGTATGAGAGTTGTGATGTCGCGACGAGAAAAGTGTGACATGTCGACACCTGTCG 2272  
Qy 781 GAAATCGCTGTTGATGACGTGGCGGGATTCACAGAGGCTTGGAACTTGGCGAAGATT 840  
Db 2271 GAAATCGCTGTTGATGACGTGGCGGGATTCACAGAGGCTTGGAACTTGGCGAAGATT 2212  
Qy 841 GAGTTGGTTCGATGAGAGCTCGGGCAATCGCGTGGTCCCGTGAATTCGTCGCGAGCT 900  
Db 2211 GAGTTGGTTCGATGAGAGCTCGGGCAATCGCGTGGTCCCGTGAATTCGTCGCGAGCT 2152  
Qy 901 GGTCTTGGGGCGAAGCTAAACAGATGCTCAGATCGTTGAGCAGAGAAATCTGCCCGGCTC 960  
Db 2151 GGTCTTGGGGCGAAGCTAAACAGATGCTCAGATCGTTGAGCAGAGAAATCTGCCCGGCTC 2092  
Qy 961 ATGTTGGGATTCATTCGCGCGGCGATCGTGAATGATTCGCACTTGGCGCGCTTAAGTC 1020  
Db 2091 ATGTTGGGATTCATTCGCGCGGCGATCGTGAATGATTCGCACTTGGCGCGCTTAAGTC 2032  
Qy 1021 TTCGCGAGATCCTCGAGCTCGTCGAAAGCTGCGCGCACTTGGGAAATCTTCGTATCAC 1080  
Db 2031 TTCGCGAGATCCTCGAGCTCGTCGAAAGCTGCGCGCACTTGGGAAATCTTCGTATCAC 1972  
Qy 1081 TTGCATTTATCGATTGCCCGCAGCGGATGTGCGGCCCGCATATATCGGTTGCAAGTGA 1140  
Db 1971 TTGCATTTATCGATTGCCCGCAGCGGATGTGCGGCCCGCATATATCGGTTGCAAGTGA 1912

RESULT 4  
US-10-007-452-5/c  
; Sequence 5, Application US/1007452  
; Publication No. US20030093701A1  
; GENERAL INFORMATION:

; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: Kostichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: CI1709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,452  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/254,868  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 5  
; LENGTH: 6334

TYPE: DNA  
ORGANISM: Rhodococcus AN12  
US-10-007-452-5

Query Match 100.0%; Score 1140; DB 14; Length 6334;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1140; Conservative 0; Indels 0; Gaps 0;

```
1 ATGACGAGGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCCCGCTCGTGTG 60
3051 ATGACGAGGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCCCGCTCGTGTG 2992

QY 61 TCCGATTAAGCGCGCATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACGCTCA 120
DB 2991 TCCGATTAAGCGCGCATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACGCTCA 2932

QY 121 GAAACATTTAAAGCTTGTGCGCGCGCGGCGGATTTCTGGCGGTAAGGCTGACCTTTGCAAC 180
DB 2931 GAAACATTTAAAGCTTGTGCGCGCGCGGCGGATTTCTGGCGGTAAGGCTGACCTTTGCAAC 2872

QY 181 GGTCCGAAAGTTCTGATTCGAGAGGCTTCCTGCTCGGAAAGGCGCTGATCTGCCCC 240
DB 2871 GGTCCGAAAGTTCTGATTCGAGAGGCTTCCTGCTCGGAAAGGCGCTGATCTGCCCC 2812

QY 241 TCGTGTGCGGAAAGTGTGCGACATCGTGCAGACGAAATTTCTCAAGTTGTCTCAT 300
DB 2811 TCGTGTGCGGAAAGTGTGCGACATCGTGCAGACGAAATTTCTCAAGTTGTCTCAT 2752

QY 301 CAACCTCGGAGCTGTGATCTGTGCGAATGTGACGATGACGATGCGGACATGACGCTGCTAG 360
DB 2751 CAACCTCGGAGCTGTGATCTGTGCGAATGTGACGATGACGATGCGGACATGACGCTGCTAG 2692

QY 361 CGGCTCCAGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 2691 CGGCTCCAGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2632

QY 421 CGTGGCGTAACGGAAGTGTGCGACATCGTGCAGACGAAATTTCTCAAGTTGTCTCAT 480
DB 2631 CGTGGCGTAACGGAAGTGTGCGACATCGTGCAGACGAAATTTCTCAAGTTGTCTCAT 2572

QY 481 ACTCAACGGAAGAAACGCGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCT 540
DB 2571 ACTCAACGGAAGAAACGCGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCT 2512

QY 541 GTGAGTGAAGAACATCTCGAATCTCTCGATGCGATGCGATGCGATGCGATGCGATGCGAT 600
DB 2511 GTGAGTGAAGAACATCTCGAATCTCTCGAATGCGATGCGATGCGATGCGATGCGATGCGAT 2452

QY 601 CTGATATCTCTGGAATTTGCTGCGCACTACGTAATTCGAGTGTCTCGATGTAAGAAAG 660
DB 2451 CTGATATCTCTGGAATTTGCTGCGCACTACGTAATTCGAGTGTCTCGATGTAAGAAAG 2392

QY 661 ATGCGGCGTGAACCTGATGCAAGTTCTGCGCGATCTGCGAATTCGCGAATTCGCGAAT 720
DB 2391 ATGCGGCGTGAACCTGATGCAAGTTCTGCGCGATCTGCGAATTCGCGAATTCGCGAAT 2332

QY 721 GGTATGAGGTTGGTATGCGACGCGAAAGTGTGCGACGATGCGACGATGCGACGATGCGAC 780
DB 2331 GGTATGAGGTTGGTATGCGACGCGAAAGTGTGCGACGATGCGACGATGCGACGATGCGAC 2272

QY 781 GAAATGCTGTGATGATGAGTGGCGGAGATTCACAGCGTTGGAATCTGTGGCGGAATTT 840
DB 2271 GAAATGCTGTGATGATGAGTGGCGGAGATTCACAGCGTTGGAATCTGTGGCGGAATTT 2212

QY 841 GAGTTGGTTCGATGAGACGTCGCGGCAATGCGCGTGTGCGCGTGTGCGCGTGTGCGCG 900
DB 2211 GAGTTGGTTCGATGAGACGTCGCGGCAATGCGCGTGTGCGCGTGTGCGCGTGTGCGCG 2152

QY 901 GGTCTTGGGCGAGAACTACAGATGCTCAAGTCTGTTAGCGAGAAAGATCTGCCCGGCTC 960
DB 2151 GGTCTTGGGCGAGAACTACAGATGCTCAAGTCTGTTAGCGAGAAAGATCTGCCCGGCTC 2092

QY 961 ATGTTGCGATCATTCGCGCGCGATCGTGTGATGATGATTCGGAATTCGCGCTTACGTC 1020
```

```
DB 2091 ATGTTGCGATCATTCGCGCGCGATCGTGTGATGATGATTCGGAATTCGCGCTTACGTC 2032
QY 1021 TTGGCGAGATCTCTGGAATCTGTGGAAGCTGGCGGCACTTTGGGAAATCTTCTGTATCAC 1080
DB 2031 TTGGCGAGATCTCTGGAATCTGTGGAAGCTGGCGGCACTTTGGGAAATCTTCTGTATCAC 1972
QY 1081 TTGATTTATGATTTGCGCGGAGCGGATGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
DB 1971 TTGATTTATGATTTGCGCGGAGCGGATGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1912
```

## RESULT 5

US-10-007-527a-7/c

Sequence 7, Application US/1007527A

Publication No. US20030044807A1

GENERAL INFORMATION:

APPLICANT: Tomb, Jean-Francois

APPLICANT: Bramucci, Michael G.

APPLICANT: Cheng, Qiong

APPLICANT: Kostichka, Kristy N.

TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors

FILE REFERENCE: C11709 US NA

CURRENT APPLICATION NUMBER: US/10/007,527A

PRIOR APPLICATION NUMBER: 60/254,868

PRIOR FILING DATE: 2000-12-12

NUMBER OF SEQ ID NOS: 30

SOFTWARE: Microsoft Office 97

SEQ ID NO 7

LENGTH: 9652

TYPE: DNA

ORGANISM: Plasmid pRHR17

US-10-007-527a-7

Query Match 100.0%; Score 1140; DB 14; Length 9652;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1 ATGACGAGGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCCCGCTCGTGTG 60
5052 ATGACGAGGTAAGTGTGTAACACCTTTCCGGCAAAAGACCGGCTCCCGCTCGTGTG 4993

QY 61 TCCGATTAAGCGCGCATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACGCTCA 120
DB 4992 TCCGATTAAGCGCGCATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACGCTCA 4933

QY 121 GAAACATTTAAAGCTTGTGCGCGCGCGGCGGATTTCTGGCGGTAAGGCTGACCTTTGCAAC 180
DB 4932 GAAACATTTAAAGCTTGTGCGCGCGCGGCGGATTTCTGGCGGTAAGGCTGACCTTTGCAAC 4873

QY 181 GGTCCGAAAGTTCTGATTCGAGAGGCTTCCTGCTCGGAAAGGCGCTGATCTGCCCC 240
DB 4872 GGTCCGAAAGTTCTGATTCGAGAGGCTTCCTGCTCGGAAAGGCGCTGATCTGCCCC 4813

QY 241 TCGTGTGCGGAAAGTGTGCGACATCGTGCAGACGAAATTTCTCAAGTTGTCTCAT 300
DB 4812 TCGTGTGCGGAAAGTGTGCGACATCGTGCAGACGAAATTTCTCAAGTTGTCTCAT 4753

QY 301 CAACCTCGGAGCTGATCTGTGCGATGATGACGATGACGATGACGATGACGATGACGATGAC 360
DB 4752 CAACCTCGGAGCTGATCTGTGCGATGATGACGATGACGATGACGATGACGATGACGATGAC 4693

QY 361 CGGCTCCAGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 4692 CGGCTCCAGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4633

QY 421 CGTGGCGTAACGGAAGTGTGCGACATCGTGCAGACGAAATTTCTCAAGTTGTCTCAT 480
DB 4632 CGTGGCGTAACGGAAGTGTGCGACATCGTGCAGACGAAATTTCTCAAGTTGTCTCAT 4573

QY 481 ACTCAACGGAAGAAACGCGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCT 540
```

Query Match	100.0%	Score 1140	DB 14	Length 9652
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1140	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
QY	1	ATGACACGCGTAAGTCTGTGAACACTTTTCGGGAAAAGACCGGCTCCCTCTCTGATGCG	60	
Db	5052	ATGACACGCGTAAGTCTGTGAACACTTTTCGGGAAAAGACCGGCTCCCTCTCTGATGCG	4993	

Db 4032 TTGGCGGAGATCTTGGACTGTGAACTGGCGGACTTTGGGAAAATCTTGATGATGAC 3973

Qy 1081 TTGCATTATCGATTGCGCGCAGCGGATGTGGCGCCCCCGATATAATTCGATTGCAAGTGA 1140

Db 3972 TTGCATTATCGATTGCGCGCAGCGGATGTGGCGCCCCCGATATAATTCGATTGCGAAGTGA 3913

RESULT 7  
US-10-007-527a-6/c  
; Sequence 6, Application US/10007527A  
; Publication No. US20030044807A1  
; GENERAL INFORMATION:  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: Kostichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: C11709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,527A  
; PRIOR FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: 60/254,868  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 6  
; LENGTH: 11241  
; TYPE: DNA  
; ORGANISM: Plasmid pRHR17  
US-10-007-527a-6

Query Match 100.0%; Score 1140; DB 14; Length 11241;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGGTAAGTGTGAACACCTTTCGCGAAGACCGGCTCCGTCCTGTCG 60  
DB 6641 ATGACGAGGTAAGTGTGAACACCTTTCGCGAAGACCGGCTCCGTCCTGTCG 6582  
QY 61 TCCGATTAAGCGCGGCAATCCGCGACGAATCGGACCCAACTTCAACAAATCACAGTCA 120  
DB 6581 TCCGATTAAGCGCGGCAATCCGCGACGAATCGGACCCAACTTCAACAAATCACAGTCA 6522  
QY 121 GAAACATTTAAGCCCTGTGGCCCGCGGATTTCTGGGGTGAACGGTGTGACATTTGTCAAC 180  
DB 6521 GAAACATTTAAGCCCTGTGGCCCGCGGATTTCTGGGGTGAACGGTGTGACATTTGTCAAC 6462  
QY 181 GGTCCGAAAGGTTCTGGAATCGGAGGCTTCTTCCGCGAAGGAGCTGATCTGCC 240  
DB 6461 GGTCCGAAAGGTTCTGGAATCGGAGGCTTCTTCCGCGAAGGAGCTGATCTGCC 6402  
QY 241 TGTGTGCGGGAAGAGTGTGCAATCGTGACAGCAAGAAATTTCTCAAGTTGTGTCAT 300  
DB 6401 TGTGTGCGGGAAGAGTGTGCAATCGTGACAGCAAGAAATTTCTCAAGTTGTGTCAT 6342  
QY 301 CAATCGGAGCTGATCTGTGGATGTGTGACATGACATGCGCCATTAAGTGTGTCAG 360  
DB 6341 CAATCGGAGCTGATCTGTGGATGTGTGACATGACATGCGCCATTAAGTGTGTCAG 6282  
QY 361 CGGCTCCAGACCTATGAGCTGAGACTTTCGCGAGCTCGAAGAGCTGACCAAGGTCGT 420  
DB 6281 CGGCTCCAGACCTATGAGCTGAGACTTTCGCGAGCTCGAAGAGCTGACCAAGGTCGT 6222  
QY 421 CGTTGGGTAAGGAACTGTAATGTACGAGCTGCGAGGATACGTGCGCTGTGAATC 480  
DB 6221 CGTTGGGTAAGGAACTGTAATGTACGAGCTGCGAGGATACGTGCGCTGTGAATC 6162  
QY 481 ACTGACGGAAGAAACGCGCTGCGACGTTCACGTTTCAAGGCTACTCATGTTCAATGAGT 540  
DB 6161 ACTGACGGAAGAAACGCGCTGCGACGTTCACGTTTCAAGGCTACTCATGTTCAATGAGT 6102  
QY 541 GTGAGTGAAGAACTCCGATCCCTTCGAGATGCGATGTTTCAATGAGTGAATTCGCA 600  
DB 6101 GTGAGTGAAGAACTCCGATCCCTTCGAGATGCGATGTTTCAATGAGTGAATTCGCA 6042  
QY 601 CTGATATCTCTGGGATTTGCTGCGCACTAAGTAATCGGAGTGTCTGATGTAAGAAAG 660  
DB 6041 CTGATATCTCTGGGATTTGCTGCGCACTAAGTAATCGGAGTGTCTGATGTAAGAAAG 5982  
QY 661 ATGCGGAGTGAAGCTGATCAAGTTCTGCGTGTGATCTGACGAAATTTGATCTGGGCTT 720

DB 5981 ATGCGGAGTGAAGCTGATCAAGTTCTGCGTGTGATCTGACGAAATTTGATCTGGGCTT 5922  
QY 721 GGTATGAGGTTGTAGTGTGCGGACGAAAGAGTGTGCAATGCGAACCTGACCTGTG 780  
DB 5921 GGTATGAGGTTGTAGTGTGCGGACGAAAGAGTGTGCAATGCGAACCTGACCTGTG 5862  
QY 781 GAAATCGGTGTGATGCGATGTGGGCGGAGTCCACAAAGGTTGAACGTGTGCGAGATTT 840  
DB 5861 GAAATCGGTGTGATGCGATGTGGGCGGAGTCCACAAAGGTTGAACGTGTGCGAGATTT 5802  
QY 841 GAGTTGTGATGAGGAGCTGCGGCAATCGCGTGTGCTCCGTAATTCGTCGCGAGCT 900  
DB 5801 GAGTTGTGATGAGGAGCTGCGGCAATCGCGTGTGCTCCGTAATTCGTCGCGAGCT 5742  
QY 901 GGTCTTGGGAGAACTTAACAGATCTCAGATCTTTAGACAGAAAGAAATTCGCGGTC 960  
DB 5741 GGTCTTGGGAGAACTTAACAGATCTCAGATCTTTAGACAGAAAGAAATTCGCGGTC 5682  
QY 961 ATGTTTGGATCATTTCCGCGCGGATCCGATGATGATGATTTGGAATTTGTCGCTTACGTC 1020  
DB 5681 ATGTTTGGATCATTTCCGCGCGGATCCGATGATGATGATTTGGAATTTGTCGCTTACGTC 5622  
QY 1021 TTCGCGAGATCTCTGCACTCGTCCGAGCTGCGCGGACTTGGGAAATCTTGTGATCAC 1080  
DB 5621 TTCGCGAGATCTCTGCACTCGTCCGAGCTGCGCGGACTTGGGAAATCTTGTGATCAC 5562  
QY 1081 TTGCATTTATCGATTGCCCGGACGAGATGTGCGGCCCGCATATATTCGTTTGCAGTCA 1140  
DB 5561 TTGCATTTATCGATTGCCCGGACGAGATGTGCGGCCCGCATATATTCGTTTGCAGTCA 5502

RESULT 8  
US-10-007-452-6/c  
; Sequence 6, Application US/10007452  
; Publication No. US20030093701A1  
; GENERAL INFORMATION:  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: Kostichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: C11709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,452  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/254,868  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 6  
; LENGTH: 11241  
; TYPE: DNA  
; ORGANISM: Plasmid pRHR17  
US-10-007-452-6

Query Match 100.0%; Score 1140; DB 14; Length 11241;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGGTAAGTGTGAACACCTTTCGCGAAGACCGGCTCCGTCCTGTCG 60  
DB 6641 ATGACGAGGTAAGTGTGAACACCTTTCGCGAAGACCGGCTCCGTCCTGTCG 6582  
QY 61 TCCGATTAAGCGCGGCAATCCGCGACGAATCGGACCCAACTTCAACAAATCACAGTCA 120  
DB 6581 TCCGATTAAGCGCGGCAATCCGCGACGAATCGGACCCAACTTCAACAAATCACAGTCA 6522  
QY 121 GAAACATTTAAGCCCTGTGGCCCGCGGATTTCTGGGGTGAACGGTGTGACATTTGTCAAC 180  
DB 6521 GAAACATTTAAGCCCTGTGGCCCGCGGATTTCTGGGGTGAACGGTGTGACATTTGTCAAC 6462  
QY 181 GGTCCGAAAGGTTCTGGAATCGGAGGCTTCTTCCGCGAAGGAGCTGATCTGCC 240





```
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94798
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-94798
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 454;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 324 GATGTGACGATGACCATGC 343
Db 266 GATGTGACGATGACCATGC 285
```

```
RESULT 12
US-10-027-632-94798
; Sequence 94798, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94798
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-94798
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 13; Length 454;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 324 GATGTGACGATGACCATGC 343
Db 266 GATGTGACGATGACCATGC 285
```

```
RESULT 13
US-10-225-810-5/c
; Sequence 5, Application US/10225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
; APPLICANT: Bermingham, Jr., John R.
; TITLE OF INVENTION: Transdorins and Methods of Using Transdorin
; FILE REFERENCE: McLaugh-07165
; CURRENT APPLICATION NUMBER: US/10/225,810
```

```
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 2055
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-225-810-5
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 2055;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 324 GATGTGACGATGACCATGC 343
Db 1054 GATGTGACGATGACCATGC 1035
```

```
RESULT 14
US-09-805-456-1/c
; Sequence 1, Application US/09805456
; Publication No. US20030170778A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01062
; CURRENT APPLICATION NUMBER: US/09/805,456
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2093
; TYPE: DNA
; ORGANISM: Human
US-09-805-456-1
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 2093;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 324 GATGTGACGATGACCATGC 343
Db 1010 GATGTGACGATGACCATGC 991
```

```
RESULT 15
US-09-805-456-3/c
; Sequence 3, Application US/09805456
; Publication No. US20030170778A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01062
; CURRENT APPLICATION NUMBER: US/09/805,456
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 46649
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(46649)
; OTHER INFORMATION: n = A,T,C or G
US-09-805-456-3
```

```
Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 46649;
```



Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GATGTGACGATGACCATGC 343

Db 32694 GATGTGACGATGACCATGC 32675

Search completed: November 8, 2003, 23:09:19  
Job time : 306 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 21:28:22 ; Search time 1954 Seconds

(without alignments)  
14179.684 Million cell updates/sec

Title: US-10-007-527a-1

Perfect score: 1140

Sequence: 1 atgaccagcgaagtcgctga.....taatatcggttcgcaagtcga 1140

Scoring table: OLIGO\_NWC

Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST.\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the chance being printed, and is derived by analysts of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	20	1.8	494	9 A1953890 wx70g08.x
2	20	1.8	538	28 A0392965 A0392965 CITR1-EI-
3	20	1.8	609	14 CA732898 CA732898 wipic.pk0
4	20	1.8	632	12 BM696484 BM696484 UI-E-DWO-

Result No.	Score	Query Match length	ID	Description
5	20	1.8	657	10 BE149632 BE149632 RCL-HT025
6	20	1.8	764	29 B2441145 B2441145 BOND555TR
7	20	1.8	767	28 BH606297 BH606297 BOG0277F
8	20	1.8	781	28 BH716628 BH716628 BOHW233TR
9	20	1.8	790	29 AL043182 AL043182 DXFZP434G
10	20	1.8	820	29 B2447351 B2447351 BONTN04TF
11	20	1.8	858	28 BH079737 BH079737 RPT-24-3
12	20	1.8	1063	28 BH722906 BH722906 BOWM275TF
13	20	1.8	1315	29 B2576656 B2576656 BM2_504
14	19	1.7	340	29 CC426700 CC426700 PUH8R89TD
15	19	1.7	376	9 AV697120 AV697120 AV697120
16	19	1.7	390	9 AA459477 AA459477 aa27b05.r
17	19	1.7	415	14 N40228 N40228 yx9yb08.r1
18	19	1.7	422	9 AW778576 AW778576 EK51911.y
19	19	1.7	423	9 AW732682 AW732682 db11b07.y
20	19	1.7	423	14 H85243 H85243 yv85d07.r1
21	19	1.7	452	9 AA201616 AA201616 LD04711.5
22	19	1.7	462	9 AJ318275 AJ318275 AJ318275
23	19	1.7	470	9 AJ412271 AJ412271 AJ318276
24	19	1.7	475	9 AJ318276 AJ318276 BQ528871
25	19	1.7	552	13 BQ528871 BQ528871 3524_1_42
26	19	1.7	560	13 BQ485604 BQ485604 3524_1_16
27	19	1.7	608	28 BH736326 BH736326 BH736326
28	19	1.7	630	9 AA193531 AA193531 z41a11.r
29	19	1.7	643	14 CD24625 CD24625 CCCT_35.C
30	19	1.7	646	12 BM679842 BM679842 UI-E-EJO-
31	19	1.7	667	28 A0566635 A0566635 HS_2105.B
32	19	1.7	834	29 B2451465 B2451465 BONGC57TF
33	19	1.7	844	29 CC011138 CC011138 PUDEB72TD
34	19	1.7	870	29 B2746327 B2746327 PUDAJ49TD
35	19	1.7	947	10 BF239778 BF239778 601906381
36	19	1.7	1013	29 CC283164 CC283164 CH261-138
37	19	1.7	1039	29 CNO20W9 CNO20W9 Tetradon
38	19	1.7	1109	13 BQ276654 BQ276654 AGENCOURT
39	18	1.6	173	14 CB858146 CB858146 RX85 Medi
40	18	1.6	242	9 AA171157 AA171157 ms43c02.r
41	18	1.6	290	10 BA415255 BA415255 BA415255
42	18	1.6	307	10 BB381929 BB381929 BB381929
43	18	1.6	320	10 BB752403 BB752403 BB752403
44	18	1.6	332	9 AA490628 AA490628 EST249182
45	18	1.6	370	10 BPS95561 BPS95561 su66f04.y

## ALIGNMENTS

RESULT 1  
A1953890  
LOCUS  
DEFINITION  
A1953890 494 bp mRNA linear EST 09-MAR-2000  
wx70g08.x1 NCI CGAP Brn53 Homo sapiens cDNA clone IMAGE:2549054 3'  
similar to SW:YXK1 YEAST P50944 HYPOTHETICAL 80.0 KD PROTEIN IN  
POLI-RAS2 INTERGENIC REGION. ; mRNA sequence.

ACCESSION  
A1953890  
VERSION  
A1953890.1 GI:5746200  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens

REFERENCE  
1 (bases 1 to 494)  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
AUTHORS  
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
NATIONAL CANCER INSTITUTE / NATIONAL INSTITUTE OF NEUROLOGICAL  
DISORDERS AND STROKE, BRAIN TUMOR GENOME ANATOMY PROJECT  
(CGAP/BRGAP), Tumor Gene Index

JOURNAL  
COMMENT  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Chris Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life  
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 1564 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence sc0p: 368.

## FEATURES

## source

1..494

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2549054"  
 /tissue\_type="three pooled meningiomas"  
 /lab\_host="DH10B"  
 /clone\_1ib="NCI CGAP Brn53"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 library constructed by Life Technologies."

## BASE COUNT

139 a 108 c 118 g 129 t

## ORIGIN

Query Match 1.8%; Score 20; DB 9; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTCGATGACCATGC 343  
 |||||  
 Db 372 GATGTCGATGACCATGC 391

## RESULT 2

## LOCUS

AO392965 538 bp DNA linear GSS 06-MAR-1999  
 CITBI-EI-2544L11.TR CITBI-EI Homo sapiens genomic clone 2544L11,  
 genomic survey sequence.

## ACCESSION

AO392965 GI:4363988

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

1 (bases 1 to 538)  
 Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and  
 Venter,J.C.  
 Use of BAC End Sequences from Caltech Libraries for Sequence-Ready  
 Map Building  
 Unpublished

## JOURNAL

## COMMENT

Other GSSs: CITBI-EI-2544L11.TF  
 Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Bukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/cdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: M13 Reverse  
 Class: BAC ends.

## FEATURES

## source

Location/Qualifiers

1..538  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="2544L11"  
 /sex="male"  
 /cell\_type="sperm"  
 /clone\_1ib="CITBI-EI"  
 /note="Vector: pBeloBAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
 Caltech Human BAC Library D"

## BASE COUNT

115 a 148 c 113 g 162 t

## ORIGIN

Query Match 1.8%; Score 20; DB 28; Length 538;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTCGATGACCATGC 343  
 |||||  
 Db 205 GATGTCGATGACCATGC 186

## RESULT 3

## LOCUS

CA732898 609 bp mRNA linear EST 26-NOV-2002  
 wlpic.pk006.a5 wlpic Triticum aestivum cDNA clone wlpic.pk006.a5 5'  
 end, mRNA sequence.

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

EST.  
 Triticum aestivum (bread wheat)  
 Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
 ; Triticeae; Triticum.  
 1 (bases 1 to 609)  
 Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,  
 Miao,G., Caraher,N. and Hanafey,M.K.  
 Dupont Wheat cDNA Sequence  
 Unpublished

## JOURNAL

## COMMENT

Contact: Scott V. Tingey  
 Crop Genetics  
 E. I. Dupont de Nemours and Company  
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
 Tel: 302-631-2602  
 Fax: 302-631-2607  
 Email: Scott.V.Tingey@USA.dupont.com  
 Seq primer: M13.

## FEATURES

## source

Location/Qualifiers

1..609  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4565"  
 /clone="wlpic.pk006.a5"  
 /tissue\_type="lemma and palea"  
 /lab\_host="DH10B"  
 /clone\_1ib="wlpic"  
 /note="Vector: pBluescript SK+; Site\_1: EcoRI; Site\_2:  
 XhoI; wheat (Triticum aestivum, H1 Line) lemma and palea"

## BASE COUNT

95 a 191 c 185 g 133 t

## ORIGIN

Query Match 1.8%; Score 20; DB 14; Length 609;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 CGATGTCGATGACCATG 342  
 |||||  
 Db 457 CGATGTCGATGACCATG 438

## RESULT 4

## LOCUS

BM696484 632 bp mRNA linear EST 28-FEB-2002  
 UI-E-DWO-agj-n-10-0-UI.r1 UI-E-DWO Homo sapiens cDNA clone  
 UI-E-DWO-agj-n-10-0-UI 5', mRNA sequence.

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 632)  
 Bernaldo,M.F., Lennon,G. and Soares,M.B.

**TITLE** Normalization and subtraction: two approaches to facilitate gene discovery

**JOURNAL** Genome Res. 6 (9), 791-806 (1996)

**MEDLINE** 97044477

**PUBMED** 8889548

**COMMENT** Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

**FEATURES** Tissue Procurement: Dr. Gregg Hageman  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1..632  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DW0-agj-n-10-0-UI"  
/issue\_type="1ens"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; site\_1: EcoR I; site\_2: Not I; UI-E-DW0 is a CDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CAGTTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

**BASE COUNT** 123 a 196 c 141 g 172 t

**ORIGIN**

**Query Match** 1.8%; Score 20; DB 12; Length 632;  
**Best Local Similarity** 100.0%; Pred.No. 28;  
**Matches** 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 324 GATGTCGACGATGACCATGC 343  
|||||  
Db 370 GATGTCGACGATGACCATGC 351

**RESULT 5** BE149632 657 bp mRNA linear EST 21-JUN-2000  
**LOCUS** BE149632  
**DEFINITION** RCL-RT0256-280300-017-d12 RT0256 Homo sapiens cDNA, mRNA sequence.  
**ACCESSION** BE149632  
**VERSION** BE149632.1 GI:8612353  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens

**REFERENCE** Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 657)  
**AUTHORS** Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,M.Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

**TITLE** Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

**MEDLINE** 20202663

**PUBMED** 10737800

**COMMENT** Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

**FEATURES** Email: asimpson@ludwig.org.br  
This sequence was derived from the PAFBSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2-RCL-HT0256-280300-017-d12&f3=2000-03-28&f4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 25  
High quality sequence stop: 637.  
Location/Qualifiers  
1..657  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="adult"  
/clone\_lib="HT0256"  
/note="Organ: head neck; Vector: puc18; site\_1: SmaI; site\_2: SmaI; A mini-library was made by cloning products derived from ORFESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

**BASE COUNT** 183 a 144 c 182 g 148 t

**ORIGIN**

**Query Match** 1.8%; Score 20; DB 10; Length 657;  
**Best Local Similarity** 100.0%; Pred.No. 29;  
**Matches** 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 324 GATGTCGACGATGACCATGC 343  
|||||  
Db 230 GATGTCGACGATGACCATGC 249

**RESULT 6** BZ441145 764 bp DNA linear GSS 13-DEC-2002  
**LOCUS** BZ441145  
**DEFINITION** BONDERS5TR BO.1.6.2\_KB\_tot Brassica oleracea genomic clone BONDERS5, genomic survey sequence.  
**ACCESSION** BZ441145  
**VERSION** BZ441145.1 GI:26698749  
**KEYWORDS** GSS.  
**SOURCE** Brassica oleracea  
**ORGANISM** Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

**REFERENCE** 1 (bases 1 to 764)  
**AUTHORS** Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.  
**JOURNAL** Whole genome shotgun sequencing of Brassica oleracea  
**COMMENT** Other GSSs: BONDERS5TF  
Contact: Chris Town  
TIGR  
5712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR

Class: sheared ends.  
 FEATURES Location/Qualifiers  
 source 1..764  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BONDR55"  
 /clone\_1lb="BO\_1.6-2\_KB\_tot"  
 /note="Vector: pHO51; Site 1: BstXI; 1.6-2 kb sheared total DNA inserted into pHO51 using BstXI linkers"  
 BASE COUNT 246 a 136 c 150 g 232 t  
 ORIGIN  
 Query Match 1.8%; Score 20; DB 29; Length 764;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 98 AACTTCACAAATCACACG 117  
 |||||  
 DB 423 AACTTCACAAATCACACG 442  
 |||||  
 RESULT 7  
 LOCUS BH606297 767 bp DNA linear GSS 15-DEC-2001  
 DEFINITION BOGJG27F BOGJ Brassica oleracea genomic clone BOGJG27, genomic survey sequence.  
 ACCESSION BH606297  
 VERSION BH606297.1 GI:17858743  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; eurosid II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 767)  
 TOWN,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.  
 AUTHORS Whole genome shotgun sequencing of Brassica oleracea  
 TITLE Unpublished  
 JOURNAL Other GSSs: BOGJG27TR  
 COMMENT Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.  
 FEATURES Location/Qualifiers  
 source 1..767  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGJG27"  
 /clone\_1lb="BOGJ"  
 /note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"  
 BASE COUNT 192 a 187 c 188 g 200 t  
 ORIGIN  
 Query Match 1.8%; Score 20; DB 28; Length 767;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 304 CTCGGAGCTGATCTGTGC 323  
 |||||  
 DB 346 CTCGGAGCTGATCTGTGC 365  
 |||||  
 RESULT 8

BH716628  
 LOCUS BH716628 781 bp DNA linear GSS 20-FEB-2002  
 DEFINITION BOHNV23TR BO\_2.3\_KB Brassica oleracea genomic clone BOHNV23, genomic survey sequence.  
 ACCESSION BH716628  
 VERSION BH716628.1 GI:18812783  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; eurosid II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 781)  
 TOWN,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.  
 AUTHORS Whole genome shotgun sequencing of Brassica oleracea  
 TITLE Unpublished  
 JOURNAL Other GSSs: BOHNV23TF  
 COMMENT Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.  
 FEATURES Location/Qualifiers  
 source 1..781  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOHNV23"  
 /clone\_1lb="BO\_2.3\_KB"  
 /note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"  
 BASE COUNT 218 a 200 c 160 g 203 t  
 ORIGIN  
 Query Match 1.8%; Score 20; DB 28; Length 781;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 304 CTCGGAGCTGATCTGTGC 323  
 |||||  
 DB 663 CTCGGAGCTGATCTGTGC 682  
 |||||  
 RESULT 9  
 LOCUS AL043182 790 bp mRNA linear EST 29-FEB-2000  
 DEFINITION DKF2p434G1123.r1.434 (synonym: htes3) Homo sapiens cDNA clone DKF2p434G1123.5', mRNA sequence.  
 ACCESSION AL043182  
 VERSION AL043182.1 GI:5422591  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 790)  
 Blum,H., Bauersachs,S., Mewes,H.W., Gassenhuber,J. and Wiemann,S.  
 AUTHORS EST (Blum, et al.)  
 TITLE Unpublished  
 JOURNAL Contact: Blum H  
 COMMENT MIPs  
 Am Klopferpitz 18a D-82152 Martinsried, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by IMU (Ludwig Maximilians University,  
 Munich/Germany) within the cDNA sequencing consortium of the German Genome Project.

No s1 sequence available.  
This clone (DKFZp434G1123) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clonezrpd.de.

## FEATURES

Location/Qualifiers

1..790

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="DKFZp434G1123"

/tissue\_type="testis"

/dev\_stage="adult"

/lab\_host="DH10B"

/clone\_1lb="434 (synonym: htes3)"

/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"

BASE COUNT 151 a 249 c 182 g 204 t 4 others

## ORIGIN

Query Match 1.8%; Score 20; DB 9; Length 790;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 324 GATGTGACGATGACCATGC 343

Db 470 GATGTGACGATGACCATGC 451

## RESULT 10

B2447351

LOCUS B2447351 820 bp DNA linear GSS 13-DEC-2002  
DEFINITION BOMF04TR\_BO.1.6\_2\_KB\_for\_Brassica\_oleracea\_genomic\_clone\_BOMF04,  
genomic survey sequence.

ACCESSION B2447351

VERSION B2447351.1 GI:26712635

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
1 (bases 1 to 820)

Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished

COMMENT Other GSSs: BOMF04TR

CONTACT: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

## FEATURES

Location/Qualifiers

1..820

/organism="Brassica oleracea"

/mol\_type="genomic DNA"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOMF04"

/clone\_1lb="BO\_1.6\_2\_KB\_tot"

/note="Vector: pHS01; Site 1: BstXI; 1.6-2 kb sheared  
total DNA inserted into pHS01 using BstXI linkers"

BASE COUNT 254 a 170 c 169 g 227 t

## ORIGIN

Query Match 1.8%; Score 20; DB 29; Length 820;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 AACTGACAAATCACCAG 117

|||||

Db 45 AACTGACAAATCACCAG 64

## RESULT 11

BH079737/c

LOCUS BH079737

DEFINITION BH079737 858 bp DNA linear GSS 18-JUL-2001  
RPCI-24-321D14.TV RPCI-24 Mus musculus genomic clone RPCI-24-321D14  
, genomic survey sequence.

ACCESSION BH079737

VERSION BH079737.1 GI:14899334

KEYWORDS GSS.

SOURCE Mus musculus

ORGANISM Mus musculus (house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other GSSs: RPCI-24-321D14.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC  
library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC  
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
page: [http://www.tigr.org/cdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html)

Plate: 321 row: D column: 14

Seq primer: T7

Class: BAC ends.

## FEATURES

Location/Qualifiers

1..858

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-24-321D14"

/sex="Male"

/cell\_type="Spleen/Brain"

/clone\_1lb="RPCI-24"

/note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;  
RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
library was cloned in the pTARBAC1 cloning vector at the  
BamHI sites using MboI partially digested male C57BL/6J  
DNA."

BASE COUNT 213 a 185 c 177 g 283 t

## ORIGIN

Query Match 1.8%; Score 20; DB 28; Length 858;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 588 GTGACTTCCAACTCGTAT 607

|||||

Db 166 GTGACTTCCAACTCGTAT 147

## RESULT 12

BH722906/c

LOCUS BH722906

DEFINITION BOMJ275TR\_BO.2.3\_KB\_Brassica\_oleracea\_genomic\_clone\_BOMJ275,  
genomic survey sequence.

ACCESSION BH722906

VERSION BH722906.1 GI:18826669

KEYWORDS GSS.

SOURCE Brassica oleracea

```

ORGANISM
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eucosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 1063)
TOWN,C.D., Van Aken,S., Utebbeck,T., Koo,H. and Fraser,C.M.
TITLE
Whole genome shotgun sequencing of Brassica oleracea
JOURNAL
Unpublished
COMMENT
Other GSSs: BOMJ275TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..1063
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone_lib="BOMJ275"
/notes="Vector: pHOs1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOs1 using BstXI linkers"
BASE COUNT
332 a 227 c 186 g 318 t
ORIGIN
Query Match 1.8%; Score 20; DB 28; Length 1063;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY
98 AACTCAACAATCACCAG 117
|||||
Db 504 AACTCAACAATCACCAG 485

RESULT 13
LOCUS B2576656 1315 bp DNA linear GSS 17-DEC-2002
DEFINITION mesh2_504.y2 msh Pseudomonas aeruginosa genomic clone msh2_504,
genomic survey sequence.
ACCESSION B2576656
VERSION B2576656.1 GI:27211717
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1315)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE
Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: Shotgun.
FEATURES
Location/Qualifiers
source
1..1315
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="M5H"
/db_xref="taxon:287"
/clone="msh2_504"

```

```

/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun
library."
BASE COUNT
359 a 352 c 221 g 383 t
ORIGIN
Query Match 1.8%; Score 20; DB 29; Length 1315;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY
227 GCTGATCTGCCCTGCTGT 246
|||||
Db 441 GCTGATCTGCCCTGCTGT 460

RESULT 14
LOCUS CC426700 340 bp DNA linear GSS 19-MAY-2003
DEFINITION PUHER897D ZM_0.6_1.0_KB Zea mays genomic clone ZMMBT434010,
genomic survey sequence.
ACCESSION CC426700
VERSION CC426700.1 GI:30906790
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 340)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utebbeck,T., Resnick
,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.
TITLE
Maize Genomes Consortium
JOURNAL Unpublished
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..340
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBT434010"
/notes="Vector: pCR4-toFO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"
BASE COUNT
118 a 72 c 74 g 76 t
ORIGIN
Query Match 1.7%; Score 19; DB 29; Length 340;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY
1030 ATCTCGAGCTGCTGAAG 1048
|||||
Db 300 ATCTCGAGCTGCTGAAG 318

RESULT 15
LOCUS AV697120 376 bp mRNA linear EST 16-JAN-2002
DEFINITION AV697120 GKC Homo sapiens cDNA clone GKCE9210 5', mRNA sequence.
ACCESSION AV697120
VERSION AV697120.1 GI:10298983
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



REFERENCE  
AUTHORS

1 (bases 1 to 376)  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

TITLE

Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,  
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,  
Shen,X., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,  
Hu,G., Gu,J., Chen,Z. and Han,Z.  
Insight into hepatocellular carcinogenesis at transcriptome level  
by comparing gene expression profiles of hepatocellular carcinoma  
with those of corresponding noncancerous liver

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
21625106  
11752456

Contact: Zeguang Han

Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
201203, P. R. China

Tel: 86-21-50801919(ex.45)

Fax: 86-21-50801922

Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

1..376  
Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="GKCE0E10"

/tissue\_type="hepatocellular carcinoma"

/dev\_stage="Adult"

/lab\_host="SOLR"

/clone\_1b="GKC"

/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:

XhoI"

BASE COUNT 76 a 113 c 113 g 74 t  
ORIGIN

Query Match

Best Local Similarity 100.0%; Score 19; DB 9; Length 376;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GCTGATCTGCCCCCTGCTG 245

Db 254 GCTGATCTGCCCCCTGCTG 272

Search completed: November 8, 2003, 23:02:25  
Job time : 1958 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 12:11:30 ; Search time 2940 Seconds  
(without alignments)

15862.917 Million cell updates/sec

Title: US-10-007-527a-1

Perfect score: 1140

Sequence: 1 atgaccagcgaatgcgcga.....taatacggcttcgcaatgca 1140

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hcg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_scs: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: gb\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_scs: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_hcg\_hum: \*  
31: em\_hcg\_inv: \*  
32: em\_hcg\_other: \*  
33: em\_hcg\_mus: \*  
34: em\_hcg\_pin: \*  
35: em\_hcg\_rod: \*  
36: em\_hcg\_mam: \*  
37: em\_hcg\_vtl: \*  
38: em\_sy: \*  
39: em\_hcgo\_hum: \*  
40: em\_hcgo\_mus: \*  
41: em\_hcgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1140	100.0	1140	6	AX548642 Sequence
2	1140	100.0	6334	6	AY178757 Rhodococc
3	1140	100.0	6334	6	AX548646 Sequence
4	1140	100.0	9652	6	AX548648 Sequence
5	1140	100.0	11241	6	AX548647 Sequence
6	1140	100.0	11241	12	AY180162 Shuttle v
7	98.4	8.3	3540	1	AY150274 Propionib
8	98.4	7.8	2439	1	APU83788 Coryneb
9	55.8	4.9	4603	1	AY172684 Coryneb
10	51.6	4.5	2297	6	E17316
11	51.6	4.5	5750	1	AF085719
12	49	4.3	125020	9	AF429315
13	47.2	4.1	2051	1	PER6662
14	46.4	4.1	125020	9	AF429315
15	44.4	3.9	2000	6	AX655393
16	42.4	3.7	9367	1	SM243257
17	41.4	3.6	13651	1	AE011844
18	41.2	3.6	12855	1	AE333885
19	40.6	3.6	1841	6	AX700509
20	40.6	3.6	1841	10	MMPTX3
21	40.6	3.6	163132	2	AC121312
22	39.2	3.4	127764	2	AC144481
23	38.8	3.4	248154	2	AC111496
24	38.8	3.4	249495	2	AC095394
25	38.4	3.4	205178	2	AC132020
26	38.4	3.4	228187	2	AC127107
27	38.4	3.4	304078	2	AC129703
28	38.2	3.4	6758	1	AY172685
29	37.6	3.3	122834	9	AC092898
30	37.4	3.3	2708	6	AX700512
31	37.4	3.3	3147	12	MMU33842
32	37.4	3.3	2708	10	AF092035
33	37.4	3.3	4447	1	AF092037
34	37.4	3.3	4447	6	AR173114
35	37.4	3.3	4447	6	AR173115
36	37.4	3.3	4447	6	BD106940
37	37.4	3.3	4447	6	BD106941
38	37.4	3.3	4447	6	BD139585
39	37.4	3.3	4447	6	BD139586
40	37.4	3.3	4457	1	BPELASM1
41	37.4	3.3	5135	12	SVE133193
42	37.4	3.3	5216	12	SVE133194
43	37.4	3.3	6241	12	AF092036
44	37.4	3.3	6272	12	AF091268
45	37.4	3.3	6398	12	AF092931

#### ALIGNMENTS

RESULT 1  
AX548642  
LOCUS AX548642 1140 bp DNA linear PAT 27-NOV-2002  
DEFINITION Sequence 1 from Patent WO02055709.  
ACCESSION AX548642  
VERSION AX548642.1 GI:25813612  
KEYWORDS  
SOURCE Rhodococcus erythropolis  
ORGANISM Rhodococcus erythropolis  
Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Nocardiaceae; Rhodococcus.  
REFERENCE  
1 Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.  
Rhodococcus cloning and expression vectors  
Patent: WO 02055709-A 1 18-JUL-2002;  
JOURNAL

Pred. No. is the number of results predicted by chance to have a

FEATURES  
 source  
 E.I. DU PONT DE NEMOURS AND COMPANY (US)  
 Location/Qualifiers  
 1..1140  
 /organism="Rhodococcus erythropolis"  
 /mol\_type="genomic DNA"  
 /strain="AN12"  
 /db\_xref="taxon:1833"

BASE COUNT 241 a 287 c 347 g 265 t

ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 1140;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-286;  
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTCTGATGACACCTTTCCGGCAAGACCGGCTCCGCTCTCGTGTG 60  
 DB 1 ATGACGAGCGTAAGTCTGATGACACCTTTCCGGCAAGACCGGCTCCGCTCTCGTGTG 60

QY 61 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAGTCA 120  
 DB 61 TCCGATTAAGCGCGGATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAGTCA 120

QY 121 GAAACATTTAAGCGCTGTGGCCGCGGATTTCTGGGTGAAGGTGTGACCATTTGTCAAC 180  
 DB 121 GAAACATTTAAGCGCTGTGGCCGCGGATTTCTGGGTGAAGGTGTGACCATTTGTCAAC 180

QY 181 GGTCCGAAAGTTCTGATTCGAGGCGCTTCGTTCCCTGGGAAAGGGCTGATTCGCCCC 240  
 DB 181 GGTCCGAAAGTTCTGATTCGAGGCGCTTCGTTCCCTGGGAAAGGGCTGATTCGCCCC 240

QY 241 TCGCTGTGCGGAAAGTGTGTCGACATCGTGCAGACGAAATTTCTCAAGTTGTGTCTAT 300  
 DB 241 TCGCTGTGCGGAAAGTGTGTCGACATCGTGCAGACGAAATTTCTCAAGTTGTGTCTAT 300

QY 301 CAATCTGGGACCTGATCTGTGTGGATGTGACAGATGACATGCGGCAATCACTGTGTG 360  
 DB 301 CAATCTGGGACCTGATCTGTGTGGATGTGACAGATGACATGCGGCAATCACTGTGTG 360

QY 361 CGGCTCCACGACCTATGACATCTTTCGCGAGCGCTGGAAGAGCTGCACCAACGCTCGT 420  
 DB 361 CGGCTCCACGACCTATGACATCTTTCGCGAGCGCTGGAAGAGCTGCACCAACGCTCGT 420

QY 421 CGTTGGCGTACGGAAGGTGAAATGTAACGCTGCGACGATACGTCGCGCTGTGAATC 480  
 DB 421 CGTTGGCGTACGGAAGGTGAAATGTAACGCTGCGACGATACGTCGCGCTGTGAATC 480

QY 481 ACTCAGGAAAAAAGCGGCTGCGACGCTTCAAGCGCTACTCATTTCTCACTGTGTGAC 540  
 DB 481 ACTCAGGAAAAAAGCGGCTGCGACGCTTCAAGCGCTACTCATTTCTCACTGTGTGAC 540

QY 541 GTGAGTGAAGACATCTCGAATCTTCTCGATGCGATGTCGATCGGTGACCTTCCAA 600  
 DB 541 GTGAGTGAAGACATCTCGAATCTTCTCGATGCGATGTCGATCGGTGACCTTCCAA 600

QY 601 CTCGATCTCTGGGATTTTCGCGCACCTACGTAATTCGAGTGTCTCATGTGACAAAG 660  
 DB 601 CTCGATCTCTGGGATTTTCGCGCACCTACGTAATTCGAGTGTCTCATGTGACAAAG 660

QY 661 ATGGGGGTGAAGCTGATCAAGTTCTCGCTGCTATCTGACGAAATTTGCACTGCGCT 720  
 DB 661 ATGGGGGTGAAGCTGATCAAGTTCTCGCTGCTATCTGACGAAATTTGCACTGCGCT 720

QY 721 GGTATGAGGTGTGTGTGCGACGAGAAAAAGTGTGCAATGCGAACCGTGCACCTCTG 780  
 DB 721 GGTATGAGGTGTGTGTGCGACGAGAAAAAGTGTGCAATGCGAACCGTGCACCTCTG 780

QY 781 GAAATGCTGTTGATGACAGTGGGCGGGAATCCAAAGCGTTGAACTGTGGCGAGATTT 840  
 DB 781 GAAATGCTGTTGATGACAGTGGGCGGGAATCCAAAGCGTTGAACTGTGGCGAGATTT 840

QY 841 GAGTTTGTTCATGAGGAGTGGGCAATGCGGTGCTCCGTGAGATTTGCTGCCAGGCT 900  
 DB 841 GAGTTTGTTCATGAGGAGTGGGCAATGCGGTGCTCCGTGAGATTTGCTGCCAGGCT 900

QY 901 GGTCTTGGGCGAGAACTPAACAGATGCTCAGATCTTGAACGAGAAATCTGCCCGGTC 960  
 DB 901 GGTCTTGGGCGAGAACTPAACAGATGCTCAGATCTTGAACGAGAAATCTGCCCGGTC 960

QY 961 ATGGTTGGATCATTTCCGGGCGGATCGTGAATGATTTGAGACTTTGTGCCCTTAAGTC 1020  
 DB 961 ATGGTTGGATCATTTCCGGGCGGATCGTGAATGATTTGAGACTTTGTGCCCTTAAGTC 1020

QY 1021 TTCGCGAGATCTCGGACTCGTGAAGCTGGCGGACCTTGGGAAATCTTGGATCAC 1080  
 DB 1021 TTCGCGAGATCTCGGACTCGTGAAGCTGGCGGACCTTGGGAAATCTTGGATCAC 1080

QY 1081 TTGCAATTAATGATTCGCCGACGAGATGTGCGGCCCGCATTAATATCGGTTGCAAGTGA 1140  
 DB 1081 TTGCAATTAATGATTCGCCGACGAGATGTGCGGCCCGCATTAATATCGGTTGCAAGTGA 1140

RESULT 2  
 AY178757/c 6334 bp DNA linear BCT 01-MAR-2003  
 LOCUS  
 DEFINITION Rhodococcus erythropolis cryptic plasmid pAN12, complete sequence.  
 ACCESSION AY178757  
 VERSION AY178757.1 GI:28628256  
 KEYWORDS  
 SOURCE  
 ORGANISM Rhodococcus erythropolis  
 Rhodococcus erythropolis  
 Bacteria; Actinobacteria; Actinomycetales;  
 Corynebacterineae; Nocardiaceae; Rhodococcus.  
 REFERENCE  
 1 (bases 1 to 6334)  
 Koestichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and  
 Cheng, Q.  
 A small cryptic plasmid from Rhodococcus erythropolis:  
 Characterization and utility for gene expression  
 TITLE  
 Unpublished  
 JOURNAL  
 2 (bases 1 to 6334)  
 Koestichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and  
 Cheng, Q.  
 Direct Submission  
 Submitted (11-NOV-2002) CRAD, E.I. Dupont de Nemours Inc.,  
 Experimental Station, Wilmington, DE 19880-0328, USA  
 JOURNAL  
 FEATURES  
 source  
 location/Qualifiers  
 1..6334  
 /organism="Rhodococcus erythropolis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1833"  
 /plasmid="pAN12"  
 /note="cryptic plasmid"  
 /complement (1912..3051)  
 /gene="rep"  
 /complement (1912..3051)  
 /gene="rep"  
 /note="rep"  
 /codon\_start=1  
 /transl\_table=11  
 /product="replication protein"  
 /protein\_id="AAO31602.1"  
 /db\_xref="GI:28628257"  
 /translation="MTSYSAEHLGSKDRPVLVSDDKGIHRLPKLOITSEFN  
 AGRISGVNCTIYNGPKSGFGILBSCKGMICPCAGKYGARADEISGVVHOL  
 GTGVAATMTTRHTRAGORLHDLMTGLSNAMKATNGRMKTRBEMTGCDYRAVEI  
 THGNKGVHVAHMLFSGDVSENILSFSDAMFRMTSKVSLGPAFLRNSGLDV  
 RKIGEADQVLAAYITKLASGVGVSGDGKSGHGRNRPWEIADVAGDPOALEL  
 WREFEGSGRRALMSRGLRABAGLGAELTDAQIVOEBSAPVMVATIPASMMIR  
 TCAVPFBEIIGLVAGATWNLRHHTLRPAADVRPIISVRK"  
 complement (4288..5178)  
 /gene="div"  
 /complement (4288..5178)  
 /gene="div"  
 /note="div"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative cell division protein"



Db 2691 CGGCTCCAGACCTATGAGCTGGACTTTGGCGGAGCTTGAGAAAGCTGCGACCAACGGTCTG 2632  
421 CGTTGGCGTACCGGAAGGTAAATGTAAGCTGGCTGCGACGATAGCGTGGCGCTGTGAATC 480  
Db 2631 CGTTGGCGTACCGGAAGGTAAATGTAAGCTGGCTGCGACGATAGCGTGGCGCTGTGAATC 2572  
Qy 481 ACTCAGGAAAAAAGCGCTGCGACGCTGCAAGCTTCAAGCGCTACTCACTGTTCACTAGTGTGAC 540  
Db 2571 ACTCAGGAAAAAAGCGCTGCGACGCTGCAAGCTTCAAGCGCTACTCACTGTTCACTAGTGTGAC 2512  
Qy 541 GTGAGTGAAGAACCTCTCGAATCTTCTCGGATGCGATGTTGATCGGTGAACTTCCAAA 600  
Db 2511 GTGAGTGAAGAACCTCTCGAATCTTCTCGGATGCGATGTTGATCGGTGAACTTCCAAA 2452  
Qy 601 CTGCTATCTCTGGGATTTGCTGGCGACATACGTAATTCGGGTGCTTCGATGTAAGAAAG 660  
Db 2451 CTGCTATCTCTGGGATTTGCTGGCGACATACGTAATTCGGGTGCTTCGATGTAAGAAAG 2392  
Qy 661 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGACGAAAAATTCGATCTGGCGTT 720  
Db 2391 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGACGAAAAATTCGATCTGGCGTT 2332  
Qy 721 GGTATGAGAGTTGGTATGAGTGGCGACGAAAAAGTGTGACATGCGAACCGTGGCACTTGG 780  
Db 2331 GGTATGAGAGTTGGTATGAGTGGCGACGAAAAAGTGTGACATGCGAACCGTGGCACTTGG 2272  
Qy 781 GAATGCGTGTGATGAGTGGCGGCGGAGATCCAAAGCGTTGAACTGTTGCGAGAAATTT 840  
Db 2271 GAATGCGTGTGATGAGTGGCGGCGGAGATCCAAAGCGTTGAACTGTTGCGAGAAATTT 2212  
Qy 841 GAGTTGGTTTGAATGAGACGTGCGGCGCAATCGGTGTGCTCCGTGATTCGTTGCCGAGCT 900  
Db 2211 GAGTTGGTTTGAATGAGACGTGCGGCGCAATCGGTGTGCTCCGTGATTCGTTGCCGAGCT 2152  
Qy 901 GGTCTTGGGCGCAAGCTAACAGATGCTCAAGTGTGTAAGCAAGAAATTTGCCCGGCTC 960  
Db 2151 GGTCTTGGGCGCAAGCTAACAGATGCTCAAGTGTGTAAGCAAGAAATTTGCCCGGCTC 2092  
Qy 961 ATGTTGGCGATCAATTCGCGGCGCATCGTGAATGATGATTTGGGACTGTGGCGCTTACGTC 1020  
Db 2091 ATGTTGGCGATCAATTCGCGGCGCATCGTGAATGATGATTTGGGACTGTGGCGCTTACGTC 2032  
Qy 1021 TTCGCGGAGATTCCTCGGACTCGTGAAGCTGCGCGCACTTGGGAAAAATCTTCGTGATCAC 1080  
Db 2031 TTCGCGGAGATTCCTCGGACTCGTGAAGCTGCGCGCACTTGGGAAAAATCTTCGTGATCAC 1972  
Qy 1081 TTGCATTATCGATTGCGCGCAAGCGATGTGGCGGCCCGCAATATATCGTTCCGAAAGTGA 1140  
Db 1971 TTGCATTATCGATTGCGCGCAAGCGATGTGGCGGCCCGCAATATATCGTTCCGAAAGTGA 1912

RESULT 4  
AX548648/c AX548648 9652 bp DNA linear PAT 27-NOV-2002  
LOCUS Sequence 7 from Patent WO02055709.  
DEFINITION AX548648  
ACCESSION AX548648  
VERSION AX548648.1 GI:25813616  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1 Shuttie vector pRHR17  
Shuttie vector pRHR17  
artificial sequences; vectors.

REFERENCE  
1 Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.  
TITLE Rhodococcus cloning and expression vectors  
JOURNAL Patent: WO 02055709-A 7 18-JUL-2002;  
E.I. DU PONT DE NEMOURS AND COMPANY (US)  
location/Qualifiers  
1. 9652

## FEATURES

SOURCE  
/organism="Shuttie vector pRHR17"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:214938"  
BASE COUNT 2097 a 2729 c 2559 g 2267 t

## ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 9652;  
Best Local Similarity 100.0%; Pred. No. 5.3e-286;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGACGAGGTGAAGTGTGAACACTTTCGCGCAAAAGACGGCTCTCCGTCTGTTGCG 60  
Db 5052 ATGACGAGGTGAAGTGTGAACACTTTCGCGCAAAAGACGGCTCTCCGTCTGTTGCG 4993  
Qy 61 TCCGATTAAGCGCGGATCCGCGACGACCGCAACCAACTTCAACAAATCAACAGTCA 120  
Db 4992 TCCGATTAAGCGCGGATCCGCGACGACCGCAACCAACTTCAACAAATCAACAGTCA 4933  
Qy 121 GAAACATTAAAGCGCTGTGCGCGCGCGATTCCTGCGGTGAACGATGTGCACTTGTCAAC 180  
Db 4932 GAAACATTAAAGCGCTGTGCGCGCGCGATTCCTGCGGTGAACGATGTGCACTTGTCAAC 4873  
Qy 181 GGTCCGAAAAGTTCTGGAATTCGAGAGCTTGTCTCTGCGAAAAGGCTGATCTGCCCC 240  
Db 4872 GGTCCGAAAAGTTCTGGAATTCGAGAGCTTGTCTCTGCGAAAAGGCTGATCTGCCCC 4813  
Qy 241 TGTCTGCGGGAAGAGTGGGTGACATGCTGACAGCAAGAAATTTCTCAAGTTGTTGCTCAT 300  
Db 4812 TGTCTGCGGGAAGAGTGGGTGACATGCTGACAGCAAGAAATTTCTCAAGTTGTTGCTCAT 4753  
Qy 301 CAACTCGGAGCTGATCTGTTGCGATGTGACAGTACCATGCGCCATACAGCTGTGCG 360  
Db 4752 CAACTCGGAGCTGATCTGTTGCGATGTGACAGTACCATGCGCCATACAGCTGTGCG 4693  
Qy 361 CGGCTTCAGACCTATGATCTGTTGCGATGTGACAGTACCATGCGCCATACAGCTGTGCG 420  
Db 4692 CGGCTTCAGACCTATGATCTGTTGCGATGTGACAGTACCATGCGCCATACAGCTGTGCG 4633  
Qy 421 CGTTGGCGTACGGAAGCTGAAATGTAAGCTGCGACGAAATTCGCGCGCTGTTGAATC 480  
Db 4632 CGTTGGCGTACGGAAGCTGAAATGTAAGCTGCGACGAAATTCGCGCGCTGTTGAATC 4573  
Qy 481 ACTCAGGAAAAAAGCGCTGCGACGCTGCAAGCTTCAAGCTTCACTGTTCACTAGTGTGAC 540  
Db 4572 ACTCAGGAAAAAAGCGCTGCGACGCTGCAAGCTTCAAGCTTCACTGTTCACTAGTGTGAC 4513  
Qy 541 GTGAGTGAAGAACCTCTCGAATCTTCTCGGATGCGATGTCATGCGTGAACCTTCCAAA 600  
Db 4512 GTGAGTGAAGAACCTCTCGAATCTTCTCGGATGCGATGTCATGCGTGAACCTTCCAAA 4453  
Qy 601 CTGCTATCTCTGGGATTTGCTGGCGCACTACGTAATTCGGGTGCTTCGATGTAAGAAAG 660  
Db 4452 CTGCTATCTCTGGGATTTGCTGGCGCACTACGTAATTCGGGTGCTTCGATGTAAGAAAG 4393  
Qy 661 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGACGAAAAATTCGATCTGGCGTT 720  
Db 4392 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGACGAAAAATTCGATCTGGCGTT 4333  
Qy 721 GGTATGAGAGTTGGTATGAGTGGCGACGAAAAAGTGTGACATGCGCAACCGTGCACCTGG 780  
Db 4332 GGTATGAGAGTTGGTATGAGTGGCGACGAAAAAGTGTGACATGCGCAACCGTGCACCTGG 4273  
Qy 781 GAATGCGTGTGATGAGTGGCGGCGGAGATCCAAAGCTTGGAACTGTGGCGAGAAATTT 840  
Db 4272 GAATGCGTGTGATGAGTGGCGGCGGAGATCCAAAGCTTGGAACTGTGGCGAGAAATTT 4213  
Qy 841 GAGTTGGTTTGAATGAGACGTGCGGCGCAATCGGTGTGCTCCGTGATTCGTTGCCGAGCT 900  
Db 4212 GAGTTGGTTTGAATGAGACGTGCGGCGCAATCGGTGTGCTCCGTGATTCGTTGCCGAGCT 4153  
Qy 901 GGTCTTGGGCGCAAGCTAACAGATGCTCAAGTGTGTAAGCAAGAAATTTGCCCGGCTC 960  
Db 4152 GGTCTTGGGCGCAAGCTAACAGATGCTCAAGTGTGTAAGCAAGAAATTTGCCCGGCTC 4093  
Qy 961 ATGTTGGCGATCAATTCGCGGCGCATCGTGAATGATGATTTGGACTGTGGCGCTTACGTC 1020  
Db 4092 ATGTTGGCGATCAATTCGCGGCGCATCGTGAATGATGATTTGGACTGTGGCGCTTACGTC 4033

QY 1021 TTCCGGGAGATCCTCGGACTGCTGCAAGCTGGGCGACATTTGGGAAAATCTTGATGATCAC 1080  
DB 4032 TTCCGGGAGATCCTCGGACTGCTGCAAGCTGGGCGACATTTGGGAAAATCTTGATGATCAC 3973  
QY 1081 TTGCATTATGATTTGCCCGAGCGGATGTGGCGGCCCGGATTAATATCGGTTGGCAAGTGA 1140  
DB 3972 TTGCATTATGATTTGCCCGAGCGGATGTGGCGGCCCGGATTAATATCGGTTGGCAAGTGA 3913

RESULT 5  
AX548647/c

LOCUS AX548647 11241 bp DNA linear PAT 27-NOV-2002  
DEFINITION Sequence 6 from Patent WO02055709.  
ACCESSION AX548647  
VERSION AX548647.1 GI:25813615

KEYWORDS Shuttle vector pRHR17  
SOURCE Shuttle vector pRHR17  
ORGANISM Artificial sequences; vectors.

REFERENCE 1  
AUTHORS Bramucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.  
TITLE Rhodococcus cloning and expression vectors  
JOURNAL Patent: WO 02055709-A 6 18-JUL-2002;  
E.I. DU PONT DE NEMOURS AND COMPANY (US)

FEATURES  
source 1..11241  
Location/Qualifiers  
/organism="Shuttle vector pRHR17"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:214938"

BASE COUNT 2517 a 3109 c 2945 g 2670 t  
ORIGIN

Query Match 100.0%; Score 1140; DB 6; Length 11241;  
Best Local Similarity 100.0%; Pred. No. 5.2e-286;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTCTGTAACACTTTCCGGCAAGACCGGCTCCGCTCGTGTGCG 60  
DB 6641 ATGACGAGCGTAAGTCTGTAACACTTTCCGGCAAGACCGGCTCCGCTCGTGTGCG 6582  
QY 61 TCCGATTAAGCGGGCATTCGGGCGACGACGCAACCCCAATCTCAAAATCCAGTCA 120  
DB 6581 TCCGATTAAGCGGGCATTCGGGCGACGACGCAACCCCAATCTCAAAATCCAGTCA 6522  
QY 121 GAAACATTTAAGCCTGTGGCCCGGCGGATTTCTGGCGTGAACGGTGTGACCATTTGTCAAC 180  
DB 6521 GAAACATTTAAGCCTGTGGCCCGGCGGATTTCTGGCGTGAACGGTGTGACCATTTGTCAAC 6462  
QY 181 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGCTCTGCGGAAAGGGCTGGATCTGCCCC 240  
DB 6461 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGCTCTGCGGAAAGGGCTGGATCTGCCCC 6402  
QY 241 TGTCTGCGGGAAGAAAGTCCGTGCAATCTGTGCAAGAAATTTCTCAAGTTGTGTGCTCAT 300  
DB 6401 TGTCTGCGGGAAGAAAGTCCGTGCAATCTGTGCAAGAAATTTCTCAAGTTGTGTGCTCAT 6342  
QY 301 CAACCTGGGAGCTGATCTGTGCGATGTGACATGACATGACATGACATGACATGACATGACATGAC 360  
DB 6341 CAACCTGGGAGCTGATCTGTGCGATGTGACATGACATGACATGACATGACATGACATGACATGAC 6282  
QY 361 CGGCTCCAGCAGCTATGAGCTGAGCTTTCGCGACGCTTGAAGAGCTGCGACCAACGAGTCTG 420  
DB 6281 CGGCTCCAGCAGCTATGAGCTGAGCTTTCGCGACGCTTGAAGAGCTGCGACCAACGAGTCTG 6222  
QY 421 CGTTGGCGGTACGGAAGCTGAATGTACGGCTGTGCAAGGATACGTCGCGCTGTGGAATC 480  
DB 6221 CGTTGGCGGTACGGAAGCTGAATGTACGGCTGTGCAAGGATACGTCGCGCTGTGGAATC 6162  
QY 481 ACTCAAGGAAAAAAGCGGCTGGGCGCTCAAGCTTCAAGCGCTACTCATGTTCACTGGGATGAC 540  
DB 6161 ACTCAAGGAAAAAAGCGGCTGGGCGCTCAAGCTTCAAGCGCTACTCATGTTCACTGGGATGAC 6102

QY 541 GTGAGTGAGAACATCTCGAATCCCTTCGAGTGGATGAGTTCGATCGGATGAGCTTCCAAA 600  
DB 6101 GTGAGTGAGAACATCTCGAATCCCTTCGAGTGGATGAGTTCGATCGGATGAGCTTCCAAA 6042  
QY 601 CTCGATCTCTGGGATTTGCTGCGGCACTAAGTATTCGAGGATGATCTGATGATGAGAAAG 660  
DB 6041 CTCGATCTCTGGGATTTGCTGCGGCACTAAGTATTCGAGGATGATCTGATGATGAGAAAG 5982  
QY 661 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGATCTGACGAGAAATTCGATCTGCGGCTT 720  
DB 5981 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGATCTGACGAGAAATTCGATCTGCGGCTT 5922  
QY 721 GGTATGAGGATTTGATGAGGCGACCGGAAAGATGTTCACATGAGGACCCGTCGCTGG 780  
DB 5921 GGTATGAGGATTTGATGAGGCGACCGGAAAGATGTTCACATGAGGACCCGTCGCTGG 5862  
QY 781 GAAATCGCTGTGATGACAGTGGGCGGAGATCCACAAGCGTTGAACTGAGGAGAAATTT 840  
DB 5861 GAAATCGCTGTGATGACAGTGGGCGGAGATCCACAAGCGTTGAACTGAGGAGAAATTT 5802  
QY 841 GAGTTGGTTGATGAGGACGTCGCGCAATTCGCGTGGTCCGATGATTTGGGCGGAGCT 900  
DB 5801 GAGTTGGTTGATGAGGACGTCGCGCAATTCGCGTGGTCCGATGATTTGGGCGGAGCT 5742  
QY 901 GGTCTTTGGGCGAAGCTAACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 960  
DB 5741 GGTCTTTGGGCGAAGCTAACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 5682  
QY 961 ATGTTGGCATCATTTCCGGCGGATCGTGGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
DB 5681 ATGTTGGCATCATTTCCGGCGGATCGTGGATGATGATGATGATGATGATGATGATGATGATGATG 5622  
QY 1021 TTCCGGGAGATCCTCGGACTGCTGCAAGCTGGGCGGCACTTTGGGAAAATCTTGATGATCAC 1080  
DB 5621 TTCCGGGAGATCCTCGGACTGCTGCAAGCTGGGCGGCACTTTGGGAAAATCTTGATGATCAC 5562  
QY 1081 TTGCATTATGATTTGCCCGAGCGGATGTGGCGGCCCGGATTAATATCGGTTGGCAAGTGA 1140  
DB 5561 TTGCATTATGATTTGCCCGAGCGGATGTGGCGGCCCGGATTAATATCGGTTGGCAAGTGA 5502

RESULT 6  
AY180162/c  
LOCUS AY180162 11241 bp DNA circular SYN 01-MAY-2003  
DEFINITION Shuttle vector pRHR17, complete sequence.  
ACCESSION AY180162  
VERSION AY180162.1 GI:30313707  
KEYWORDS  
SOURCE Shuttle vector pRHR17  
ORGANISM Shuttle vector pRHR17  
REFERENCES 1 (bases 1 to 11241)  
AUTHORS Kostichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarsajan, V. and Cheng, Q.  
TITLE A small cryptic plasmid from Rhodococcus erythropolis: characterization and utility for gene expression  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 11241)  
AUTHORS Kostichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarsajan, V. and Cheng, Q.  
TITLE Direct Submision  
JOURNAL Submitted (14-NOV-2002) CRD, E. I. Dupont de Nemours Inc., Experimental Station, Wilmington, DE 19880-0328, USA  
FEATURES  
source 1..11241  
Location/Qualifiers  
/organism="Shuttle vector pRHR17"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:214938"  
/focus  
/note="Rhodococcus - Escherichia coli shuttle vector; Sep I linearize pAN12 cloned into the PvuII site of pBR328"  
join(9925..11241,1..350)  
/organism="Cloning vector pBR328"

```
source
    /mol_type="genomic DNA"
    /db_xref="taxon:221946"
    3591. .9924
    /organism="Rhodococcus erythropolis"
    /mol_type="genomic DNA"
    /db_xref="taxon:1833"
    /plasmid="pAN12"
    /note="cryptic plasmid"

BASE COUNT      2517 a      3109 c      2945 g      2670 t

ORIGIN
Query Match      100.0%; Score 1140; DB 12; Length 11241;
Best Local Similarity 100.0%; Pred. No. 5.2e-286;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ATATACCAAGCTAGTGTCTGAACACCTTTCCGGAAGACCGGCTCCGCTCCGCTGTCG 60
DB      6641 ATATACCAAGCTAGTGTCTGAACACCTTTCCGGAAGACCGGCTCCGCTCCGCTGTCG 6582

QY      61  TCCGATTAAGCGCGGCAATCCGCGACGAATGCGACCCAACTTCAACAAATCAACAGTCA 120
DB      6581 TCCGATTAAGCGCGGCAATCCGCGACGAATGCGACCCAACTTCAACAAATCAACAGTCA 6522

QY      121  GAAACATTTAACCGCTGTGGCCGCGGATTTCTGCGGTGAACCGGTGTACATTGTCAAC 180
DB      6521 GAAACATTTAACCGCTGTGGCCGCGGATTTCTGCGGTGAACCGGTGTACATTGTCAAC 6462

QY      181  GGTCGCAAAAGTTCTGATTCGGAAGCGCTTCCTCGCGGAAGGCGTGGATCTGCCCC 240
DB      6461 GGTCGCAAAAGTTCTGATTCGGAAGCGCTTCCTCGCGGAAGGCGTGGATCTGCCCC 6402

QY      241  TGCTGTGCGGGAAGTCGCTGACATCTGTGACAGCAATTTCTCAAGTTGTGTCTCAT 300
DB      6401 TGCTGTGCGGGAAGTCGCTGACATCTGTGACAGCAATTTCTCAAGTTGTGTCTCAT 6342

QY      301  CAACTCGGGACTGATCTGTTCGATGTGACATGACATCAATGCGCATACAGCTGTGAC 360
DB      6341 CAACTCGGGACTGATCTGTTCGATGTGACATGACATCAATGCGCATACAGCTGTGAC 6282

QY      361  CGGCTCCAGCACTTATGATCTGACATTTCCGCGACCTCGGAACCTGGAACCAACGTCGT 420
DB      6281 CGGCTCCAGCACTTATGATCTGACATTTCCGCGACCTCGGAACCTGGAACCAACGTCGT 6222

QY      421  CGTTGGCGTACGGAACGTGAATGTATCGCTGACAGCGATACGTCGCGCTGTGAATC 480
DB      6221 CGTTGGCGTACGGAACGTGAATGTATCGCTGACAGCGATACGTCGCGCTGTGAATC 6162

QY      481  ACTCACGGAAGAAACCGCTGCGACGTCACGTTCAACGCGTACTCATGTTCAGTGTGAC 540
DB      6161 ACTCACGGAAGAAACCGCTGCGACGTCACGTTCAACGCGTACTCATGTTCAGTGTGAC 6102

QY      541  GTGATGAGAAACATCTCTCGAATCCTTCTCGGATGCGGATGTTTCCATCGGTGACCTTCCAA 600
DB      6101 GTGATGAGAAACATCTCTCGAATCCTTCTCGGATGCGGATGTTTCCATCGGTGACCTTCCAA 6042

QY      601  CTGATATCTCTGGGATTTGCTGCGCACTACGTAATCGGGTGTCTCGATGTACGAAG 660
DB      6041 CTGATATCTCTGGGATTTGCTGCGCACTACGTAATCGGGTGTCTCGATGTACGAAG 5982

QY      661  ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGACGAATTCATCTGCGCTT 720
DB      5981 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGACGAATTCATCTGCGCTT 5922

QY      721  GGATATGAGGTTGTGTATGTGGGACGGAAGAAAGTGTGACATGCGCAACCGTGCACCTGG 780
DB      5921 GGATATGAGGTTGTGTATGTGGGACGGAAGAAAGTGTGACATGCGCAACCGTGCACCTGG 5862

QY      781  GAAATCGCTGTGATGACGTGGCGGAGATCCAAAGGCTGGAATCTGGCGGAATTT 840
DB      5861 GAAATCGCTGTGATGACGTGGCGGAGATCCAAAGGCTGGAATCTGGCGGAATTT 5802

QY      841  GAGTTTGCTTCGATGGGACGTCGCGCAATCGCGGTGTCCTCGTGAATTCGTCGCCGAGCT 900
DB      5802 GAGTTTGCTTCGATGGGACGTCGCGCAATCGCGGTGTCCTCGTGAATTCGTCGCCGAGCT 900
```

```
DB      5801 GAGTTTGCTTCGATGGGACGTCGCGCAATCGCGGTGTCCTCGTGAATTCGTCGCCGAGCT 5742
QY      901  GGTCTTGGGAGAACTAACAGATGCTCAGATCTGTTAGACAGAAAGAACTCCCGGCTC 960
DB      5741 GGTCTTGGGAGAACTAACAGATGCTCAGATCTGTTAGACAGAAAGAACTCCCGGCTC 5682
QY      961  ATGTTGGGATCATTTCCGCGCGGATCGTGAATGATGATTCGACCTTGTCCGCTTACGTC 1020
DB      5681 ATGTTGGGATCATTTCCGCGCGGATCGTGAATGATGATTCGACCTTGTCCGCTTACGTC 5622
QY      1021  TTCGCGGATCCCTCGGACTGTCGAACTGCGCGGACCTTGGGAAATCTCGTATAC 1080
DB      5621 TTCGCGGATCCCTCGGACTGTCGAACTGCGCGGACCTTGGGAAATCTCGTATAC 5562
QY      1081  TTGCAATTCGATTCGCGCGGACGAGATGTCGCGCGGATATATATCGGTTTGCAGTGA 1140
DB      5561 TTGCAATTCGATTCGCGCGGACGAGATGTCGCGCGGATATATATCGGTTTGCAGTGA 5502

RESULT 7
AY150274
LOCUS      3540 bp      DNA      circular BCT 02-JAN-2003
DEFINITION Propionibacterium granulosum cryptic plasmid pPG01, complete
ACCESSION      AY150274
VERSION      AY150274.1
KEYWORDS      GI:27465054
SOURCE      Propionibacterium granulosum
ORGANISM      Propionibacterium granulosum
REFERENCE      Farrar, M.D. and Holland, K.T.
AUTHORS      Farrar, M.D. and Holland, K.T.
TITLE      Isolation and characterisation of a cryptic plasmid from the human
            skin commensal propionibacterium granulosum
JOURNAL      Unpublished
PUBDATE      2 (bases 1 to 3540)
AUTHORS      Farrar, M.D. and Holland, K.T.
TITLE      Direct Submision
JOURNAL      Submitted (13-SEP-2002) Skin Research Centre, Division of
            Microbiology, University of Leeds, Leeds LS2 9JT, UK
FEATURES
            source
            1..3540
                /organism="Propionibacterium granulosum"
                /mol_type="genomic DNA"
                /strain="pP283"
                /db_xref="taxon:33011"
                /plasmid="cryptic plasmid pPG01"
                74..280
                    /codon_start=1
                    /transl_table=11
                    /product="orf3"
                    /protein_id="AA078121.1"
                    /db_xref="GI:27465055"
                    /translation="MGQYGIILNRGYNSSHIEITIGYPMKRGVIVYMPDALLSAMPL
                        SSGFSPGHVBPRAVEYFHCADHD"
                    449..1792
                        /gene="tra"
                        449..1792
                            /gene="tra"
                            /note="plasmid transfer protein"
                            /codon_start=1
                            /transl_table=11
                            /product="TetA"
                            /protein_id="AA078122.1"
                            /db_xref="GI:27465056"
                            /translation="MAGERTVLTGYVMRLRILPVLLIGVAFACWCEPVHLVYVIGRA
                                VALAVEAPPPVQVQKMOVPARYVMWMSLAARFIESYGGDRIYRGLQWI
                                RSDRVLVLRVPAGLADSAALIERGAALIORLGRKSMRTCIYVAPAHGDIIR
                                DATAGEELPAPETSMNVVPGVPGSEVVLDSHPSHLVSGTRSGSGSFYGLL
                                DQWRHLPVYVAGDPPGILLFNEIGDMDGADLAEKRIITNDADAAYOVU.SMTDEND
                                RYIYLNCEHRDKMSRNPDESPPGRLLIYITIEYFGLIRLQNPDSARARSDDRA
                                SKAAGLVRIAYEGATVGVLLVLTQRPDAKTIIGPLRAQLTTRVTFPAQSDGLRMSH
```



gene	1979. .3280	PELSESEVKNQSNWASGVGFTEADSVILPTRFRSRYRABLTDLHRPASVQGLDILQ"
CDS	1979. .3280	/gene="rep" 1979. .3280 /gene="rep" /note="Plasmid replication protein" /codon_start=1 /transl_table=1 /product="Rep" /protein_id="AA078123.1" /db_xref="GI:27465057" /translation="MLNIVTLISKAQPHDLRCSHEPCABARSERORARGLGTTAT YAGTSDSCRPKREGRRRHYEMRODGLNPOYMPLERYKCAVVSQRIALMAHGGAG FVATDSCGFWACVCAKYSARRBDELAIVVQVAVGLGKVSMLTITQHHAGODLA ELWASLOSQWMAVTSGRWQEPFCAQLQGVQWVAVEYTHSHGHHVHVLVLSKDOP TSVDGTLIRHRKQGRRTTPPEVQOREPDEIAERMSGLKRGVDFIAGSGGLDWQTA DSGEELAEGLVYAKMSSVDGLANEATLGFAGFKARNGRNPFPOLIEDFLDTGETDLR LMRYVYASHGKRLATWSKGLRDAGHSEMSBOVAAODCCGBAVALFPHDMRQIR TAGAFLFIDLELHAGSRGVYAMLKRRIHIEIPLVPMSIST"
BASE COUNT	693 a 967 c 1066 g 814 t	
ORIGIN		
Query Match	8.3%; Score 94.4; DB 1; Length 3540;	
Best Local Similarity	56.2%; Pred. No. 1.8e-13;	
Matches 198; Conservative	0; Mismatches 151; Indels 3; Gaps 1;	
Dy	181 GGTCCGAAAGTTCTGGAATTCGAGAGCCCTTCCTTCGCGAAAGAGCTGAGATCGCCC	240
Dg	2267 GGTGATGATGGTGCCGGTTATGCGGGTTTGAGCGACGTGCGGAAGTGTTGGCTTGCCCT	2326
Dy	241 TGTCTGTCGGGAAAGTCGTCGCACTCTGCGACACGAAATTTCTCAAGTTGTTGCTCAT	300
Db	2327 GTCGTGCGGGAAGATTTCCGCGCACCGTCGTATAGCGTGGCCGATGTTGTCACAGTT	2386
Dy	301 CAATCGGGAATGAGT---CTGTTGCGATGGTGACATGACATGACGCGCATACAGCTGT	357
Db	2387 GCGGTTGACTCGGCTTCAAGGTGTCAATGCTGACGCTTATCAAGTCATCATGCTGTGT	2446
Dy	358 CAGCGAGCTCCACGACCTATGAGCTGGAATTTTCGCGACGCTTGAAGAGTCGACCAACGCT	417
Db	2447 CAGATCTCGCGGAGCTGTGGGCGATCGCTCCAGTCGGGTTGGATAGCTGTCAAGAGTGT	2506
Dy	418 CCGTGTGGCGTACGGAACGTGAATGTAACGCTGCGACGGAATACGTGCGGCTGTTGAA	477
Db	2507 CGACGGTGCGCAGGAATTTTGGCGCTCAGGTCGGCTCCAGGGATGGGTCAAGGCGAGTTGAA	2566
Dy	478 ATCTCTCACGGAAAAACGGCTGCGACGTCACGCTTACCGGCTACACATGT	529
Db	2567 GTCAACCCATGGGTGCGCATGGGTGGACGTTTACGTTGACGTTCTGTATCTT	2618

RESULT 8

APU83788

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

TITLE

AUTHORS

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

2439 bp DNA

circular BCT 15--JUN-1998

Actinomyces pyogenes plasmid pAP1, complete plasmid sequence.

U83788

U83788.1 GI:1805288

.

Arcanobacterium pyogenes

Arcanobacterium pyogenes

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Actinomycetaceae; Actinomycetaceae; Arcanobacterium.

1 (bases 1 to 2439)

Ballington, S.J., Jost, B.H. and Songer, J.G.

The Arcanobacterium (Actinomyces) pyogenes plasmid pAP1 is a member

of the pJ101/pV1 family of rolling circle replication plasmids

J. Bacteriol. 180 (12), 3233-3236 (1998)

98292760

9620977

2 (bases 1 to 2439)

Ballington, S.J., Jost, B.H. and Songer, J.G.

Direct Submission

Submitted (06-JUN-1997) Veterinary Science, University of Arizona,

FEATURES	1117 East Lowell Street, Tucson, AZ 85721, USA
source	Location/Qualifiers
	1. .2439
	/organism="Arcanobacterium pyogenes"
	/mol_type="genomic DNA"
	/strain="BBR1"
	/db_xref="taxon:1661"
	/plasmid="pAP1"
rep_origin	1. .110
	/note="single stranded origin; putative"
rep_origin	39. .44
	/note="site of single strand initiation; putative"
gene	221. .1600
	/gene="rep"
	221. .1600
	/gene="rep"
	/function="replication protein"
	/codon_start=1
	/transl_table=1
CDS	/product="Rep"
	/protein_id="AAC46399.1"
	/db_xref="GI:1805289"
	/translation="MNNLSERTALSLPAPOLIKYVIAAGGRSLKSPFGMTATYARSRGQ ASGSRSDRKSQIPBSRRGRSATHPLGNTVLTFFPSNSKTKTKRSRRSRYLRGQ LAEISTESVRKCGRVFAPLVSIDRASDRGAGYGGALHTGCVMACPVCSAKIAARR KTDLOQVDAVAVKRGMTVSMLTLTORHKKQQLKMLDALSTAMNLTSGRPMTEFEK QFGVLVAVRANEITGKRGGMVHSHVLIISEKPLSTFYGRKGRRLPYPEIYMY GSPIFAERMEAGLAKHGVDLSDGGIDMTVAADARAIGNVSMQTSPLAISEVTLL GSPFKANGRTPFOILLADILSLGDVVDLKLMEYEVASFGRRALITWSKGLRMANIG VSGEDRIASEEIGDEAIALFTHDAMQVRRFGAELLDTYESGGRAAAATKMLDFREI DMSLPPIIE"
RBS	1673. .1676
CDS	1685. .2023
	/note="ORF12"
	/codon_start=1
	/transl_table=1
	/product="unknown"
	/protein_id="AAC46400.1"
	/db_xref="GI:1805290"
	/translation="MFFIGKRTKFOPLAPCKPSPIFDGNGKRTGERSSEKRTQALITG EDGLVELTSIPQFQGLIDDAVKSRELBREKYNLAVLRNGGISTYLLTSGAHHV PAPVVEVSE"
RBS	2010. .2013
CDS	2020. .2409
	/note="ORF129"
	/codon_start=1
	/transl_table=1
	/product="unknown"
	/protein_id="AAC46401.1"
	/db_xref="GI:1805291"
	/translation="IMSESPGISNNAELQVFSRPAKREARLYMSRLIKELSEGKSHS TPTALESCEKLYMSVEPRLVDVLEHVEGRCRAGNTNSAARSLSSAPFIPAFVYGS LENGRRPAOLLTDQVAGRKTLGTG"
	2378. .2406
rep_origin	/note="double stranded origin; putative"
BASE COUNT	603 a 657 c 656 g 523 t
ORIGIN	
Query Match	7.8% Score 88.8; DB 1; Length 2439;
Best Local Similarity	55.5% Pred. No. 5.1e-12;
Matches 193; Conservative	0; Mismatches 152; Indels 3; Gaps 14;
Db	187 AAAGTTTCGATTCGAGGCGCTTCGTTCTCGCGAAGGCGTGGATCGCCCTGCTGT 246
Db	617 AAAGCCCGGATATGATGTTGTTGCAACCTTGTGAAAGCGTCTGGGCGTCCCACTCTGT 676
Db	247 GCGGAAAAGTCGGTGCATCGTCGACAGCAAAAATTTCTCAAGTTGTTGTCATCACTC 306
Db	677 AGCGGAAAATTCGCGCGCTCGCGAAAAACGACCTCCAAAGGTCTGTGACACAGCCGTA 736
Qy	307 GGAAGTGA--TCTGTTGCATGTGTGACATGACATGACATGCCCATACAGCTGTCTACGGG 363
Db	737 AAACCGGAATGACCGCTCTCAATCTTAAGCTCAACCGACCGTCAACCAAGGACAAAGG 796

Qy	364	TTCAACACCCATAGACTGGACTTTGGCAGCTGGAAAGCTGCACCAAGGTGTGT	423
Db	797	CTAAACACCTCTGGAGCCTCTTGCAAGGATGAATTCGGTTACTCTGTGCTGTGT	856
Qy	424	TGGCGTACGAAACGTGAATGTAACGGCTGCACAGGATACGTGCGCGCTGTTGAATCACT	483
Db	857	TGGATTGAGTTCAAGAGCAATTGGTTTGAATGCTGGTTATGTTCAAGCAATGAATAATCT	916
Qy	484	CACGGAAAAACGGCTGGCAGCTGCACGTTCAACGGCTACATCATCTTC	531
Db	917	CATGGAAACACCGCTGGCAGTGTGATTCACATGTTTGATTAATTTCC	964

RESULT 9	AY172684/c	AY172684	4603 bp	DNA	circular BCT 12-JAN-2003
LOCUS					
DEFINITION		Corynebacterium glutamicum plasmid pAG3, complete sequence.			
ACCESSION		AY172684			
VERSION		AY172684.1	GI:27657761		
KEYWORDS					
SOURCE					
ORGANISM		Corynebacterium glutamicum			
		Corynebacterium glutamicum			

REFERENCE	AUTHORS	JOURNAL	REFERENCE	AUTHORS	JOURNAL
1 (pages 1 to 4603)	Tauch, A., Puehler, A., Kalinowski, J. and Thierbach, G.	Plasmids in <i>Corynebacterium glutamicum</i> and their molecular classification by comparative genomics	2 (bases 1 to 4603)	Tauch, A., Puehler, A., Kalinowski, J. and Thierbach, G.	Direct Submission
Submitted (05-NOV-2002) Department of Genetics, University of					
Bielefeld, Universitaetsstrasse 25, Bielefeld D-33615, Germany					

FEATURES	Location/Qualifiers
source	1. .4603

```
source
1..4603
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/strain="22220"
/db_xref="taxon:1718"
/plasmid="pAG3"
complement(200..397)
CDS
```

```

/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="AAO18193.1"
/db_xref="GI:27657762"
/transl_table="MSK0K1SPFRAKTLAAVDREVTYPQVAHKIGTSRQAGSLMAL
HKANLIDRTFRGIYAPLEVDY"
complement(613..2061)
/gene="repa"
complement(613..2061)
CDS

```

gene complement(613..2061)  
/gene="repa"  
complement(613..2061)  
/gene="repa"  
/note="replication protein; essential for pAg3  
replication"  
/codon\_start=1  
/transl\_table=1  
/product="Repa"  
/protein\_id="AAO18194.1"  
/db\_xref="GI:27657763"  
/translation="MSILSSKQAEHATSAPNQPAPHTTGRLEGTDTTTHKKYKALKALS  
PALFPAKIMLDLQRTMAKVTEDTLIAGCGRWAGSVAISLVQAGRWGLGNHSHS  
VWGSVAIVSSICSPRANVEDAIQAMNAGRTIPAHYGVSTSPDRKORVSNPVVV  
BEGISITLTHHNSKOSTIEWDIALACQOAVNTTAMRGAGTAPGKSGYGLAHRW  
RALIETTHKMGHVLHVLPHFDVLSDDEBDSLADREDMAKAVAILMGARSPH  
GIDVVHYAASSPDASKISGTYCTKMLSGILAAFTTGTGTAKDGNPPRPIIDLDLGR  
QITKEDHAIEMBEKSGSKGRRTGWSQGRVDVLGINSDDIIDSLETDNNGSIVK  
VGRGMAVIASTTEKRAVILDAVVAATSDAEGKRVKARVDVLCFVAHTVNIIGAKTID  
LDTVPFHTIDVFKRELHEAVWSGRKOTALPVL"  
2396..2785

```

/Note="orf3"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="AAO18195.1"
/db_xref="GI:27657764"
/translation="MYRSATCLALHALNVVRCFVNSRETLMDTYATLNRICKMAT
KAELEQQLKMDQALKEQQLAFARAELEEQNLKVVYTVAVPQVVRDEMRLALAI
QGMITQCPAAAFAGMTDELESDQDH"
complement(2840..3322)

```

```

//codon_start=1
//evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="AA018196.1"
/db_xref="GI:27657765"
/translation="MTKPVNFSANDIEEAMIGVIANNNYCTDROAKTKAKIATA
LNKADELRLRGKKDLAHQDETTHILERYLRMLPEFQKALTDNNQLDLBVSRLQSLVYR
IEASEEFLRYGNGWEEVQLAEFDFTVE"
complement(3344..4030)

```

```

/codon_start=1
/evidence=not experimental
/translation_table=11
/product="hypothetical protein"
/protein_id="AAO18197.1"
/db_xref="ncgi:27657765"
/translation="MATTPSKRPFGRLAQLEKRVSSLSLEIVASNNVAHHDLSDDGKNNNA
WRYVAPKHTYIAOVEVALETTISWSDIAFVADAKALPTITDTSKLVEMELQRII
NNGIEPDSGDIYKLVTSHPSPRTALIELLEAGRGSDENWGSIILEKTSVAALTE
NMIIQARINSYFGVNLRTMYALDDRKALFNHWMNVTRSDADYDMEVPVTFVFPFKR
TNAETVYRAS"

```

gene	complement (4112. .4519)
gene	/gene="orf6"
CDS	complement (4112. .4519)
	/gene="orf6"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="AAO18198.1"
	/db_xref="GI:27657767"
	/translation="MGRNATAMFQVSTROIINISKALINIDWGDHVAKANASABEQI OSATLEADKPKRIASLEIDSVLKTCTTGERDDQIMIAISDKFLKLTSTGAPQ EVAQADAMALIDRFMTATKRAAHGSGFDNPDA"
BASE COUNT	1077 a 1276 c 1112 g 1138 t
ORIGIN	

	Query Match	Similarity	4.9%;	Score 55.8;	DB 1;	Length 4603;
	Best Local	Similarity	54.0%;	Pred. No. 0.002;	Mismatches 114;	Conservative 0;
QY	445	TACGCTTCGACGAGATACGTGCGCGCTGTGAAATCACTCAACGAAAAACGGCTGGCAC	504			
DB	1428	TATGGCAATTCGCCATTGGTACCGGAGCATTTAGAGTCAACCCACGGCAAAATGCTGGGCAT	1365			
QY	505	GTCACACGTTCACGGCTACTCATGTTCAGTGTGTACGATCGAATGAGAAATCTCGAATCC	564			
DB	1368	GTTACCTCGCAGTGTGTGTTCACGATGCGGGGTGTAAGTGTGATGAAAGAAATTC	1307			
QY	565	TTCTCGAGTGGAGATTTGATCGATCGGTGACCTCCAACTCGATCTCTTGAGATTGCTCG	624			
DB	1308	TTGGTGATTCGTGTTCGATAGAGTGGCTCTGTAAGCTGTTCGCTTGTGATGTGCTGA	1245			
QY	625	CCACTACGTAATTCGGGTGTCTCGATGTAC	655			
DB	1248	CGATCTCGTATCGCGGATGATGATGTTCG	1218			

RESULT 10  
E17316

REFERENCE	1	(bases 1 to 5750)
AUTHORS	O'Riordan, K., and Fitzgerald, G.F.	
TITLE	Molecular characterisation of a 5.75-kb cryptic plasmid from <i>Bifidobacterium breve</i> NCFB 2258 and determination of mode of replication	
JOURNAL	FEMS Microbiol. Lett. 174 (2), 285-294 (1999)	
MEDLINE	99271179	
PUBMED	10339821	
AUTHORS	O'Riordan, K. and Fitzgerald, G.F.	
TITLE	Direct Submission	
JOURNAL	Submitted (22-AUG-1998) Microbiology, University College Cork, Western Road, Cork, Ireland	
FEATURES	Location/Qualifiers	
source	1. 5750	
gene	/organism="Bifidobacterium breve"	
	/mol_type="genomic DNA"	
	/strain="NCFB 2258"	
	/db_xref="taxon:1685"	
	/plasmid="pCIBb1"	
	1463. .2062	
	/gene="par"	
	/note="partitioning gene"	
	1463. .2062	
	/gene="par"	
	/note="Par protein; putative involvement in plasmid partition"	
	/codon_start=1	
	/transl_table=11	
	/product="partitioning protein"	
	/protein_id="AAD34708.1"	
	/db_xref="GI:4972587"	
	/translation="MISTIAKSGGKTSIFLACAHATDPSLSITLADAPQGS	
	ATDMAEBAQNDPLPFWKTVAGESDIEIKXNEGLTITDTPGFTDLTOTTIKSDL	
	CIITTSFQIDIRAKWKTPEALQRAQGSATLITRANTRTKDFRMIDDLANDDANNG	
	ILFDVTRISQSTRYSKPEGHTPKRTDYADAWOEIOKAG"	
	complement (2373. .3515)	
	/gene="rep"	
	/note="replication gene"	
	complement (2373. .3515)	
	/gene="rep"	
	/note="Rep; similar to replication proteins of streptomycetes plasmids and other members of the PC194 family that replicate by a rolling circle mechanism"	
	/codon_start=1	
	/transl_table=11	
	/product="replication protein"	
	/protein_id="AAD34709.1"	
	/db_xref="GI:4972588"	
	/translation="MTNGEKARFTGMLCGSIWACPSCALIRHRAHVALAIGNHA	
	EIKRKRAADQMAEHRGQRLPPELWMSDSFGNYITRGLTLRHDRTMPLAMTLDAIKG	
	QMTKNGSPQKASERKTRIGFVRALEITTVGNQMHPIHFWPLADGDLDDQGRAMQ	
	QMLLDWKTWVRVAAYVKKQGNPYNVAPNDHGHDLDFKSGKADGAAAYITKIQ	
	GDGQVTLAQELIARQDINKRMGSVNPFLDLSGCLGSLDFQREDLMETWQATLRR	
	CIWSSGLGLEDMEVELEIDBELEIAEKADELPGHVGWVYPMRVYDKIRKSAPELTALAD	
	AAEREDQVAVRLPGGVILITLPEQDAIADGAKPGDYLPWVSVMV"	
	complement (3926. .3934)	
	/note="region conforming to double stranded origin; nicking site of rolling circle plasmids"	
	complement (3985. .4845)	
	/gene="ftsK gene homolog"	
	complement (3985. .4845)	
	/gene="ftsK gene homolog"	
	/note="hypothetical protein; similar to transfer proteins and ftsK homologs"	
	/codon_start=1	
	/transl_table=11	
	/product="unknown"	
	/protein_id="AAD34710.1"	
	/db_xref="GI:4972589"	
	/translation="MKPMWNTKPEPLPFLRPPDPATGITVKGKLANDATLIVT	
	NTSGVVGGIIPSGTAGNMGVILVALYLSGCCNIVHIDKGGDDWGSSEHATTVPRG	

DLDTVHDTLLRIIDEMKCRILASRORYSANYWNPDPKREPLEVITIIIDECOSFENAK  
GILGKRPADKXAEITTAATEIVYKRSRSGFLIPATQKPTTDSIPALRENCERIC  
FRVKTEPARAVAGDMPDGSFSPPTDIPARRGAIITIGATGDMCRAYVSEBAER  
AVANHKGVIT"

BASE COUNT 1278 a 1656 c 1616 g 1200 t  
Query Match 4.5%; Score 51.6; DB 1; Length 5750;  
Best Local Similarity 52.9%; Pred. No. 0.025;  
Matches 111; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 394 GCCTGGAAAGCTGCCGACCAAGCGTCTGCTGTCGTCGTAACGGAACGTAAATGTAAGCGCTGC 453  
Db 3212 GCGTGGACGAAGATGATTAACGGAAGCCCTTGCCACGCGCCCTCGAAGCCTTGGAATTC 3153  
QY 454 GACGATACGTGGCGCGCTGTTGAATCACTACGGAATAAGCGCTGACGCTTCACGTT 513  
Db 3152 AGGGGTTTGTCCGGCGCATGTAATCACTACGGAATGTAAGCGCTGACGCTTCACGTT 3093  
QY 514 CACGCGCTACTCATGTTCAAGTGTGACGTAAGTGAACAATCTCGAATCTTCTCGCAT 573  
Db 3092 CATTCGTCATGTTCTTCGATGCGCATCTGACGATGGCGAGCGTGAAGCAATGACAGCAA 3033  
QY 574 GCGATGTGATGCGTGGGACCTTCCAAATC 603  
Db 3032 TGGCTGCTGATGCTGTAACCAATGATGTC 3003

RESULT 12  
AF429315/c 125020 bp DNA linear PRI 18-JAN-2002  
LOCUS Homo sapiens junctophillin 3 (JPH3) gene, partial cds.  
DEFINITION AF429315  
ACCESSION AF429315.1 GI:17646244  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 125020)

REFERENCE  
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,  
Inglesoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,  
Potter,N.T., Ross,C.A. and Margolis,R.L.  
A repeat expansion in the gene encoding junctophillin-3 is  
associated with Huntington disease-like 2  
Nat. Genet. 29 (4), 377-378 (2001)  
JOURNAL  
MEDLINE 21583737  
PUBMED 11694876  
REFERENCE 2 (bases 1 to 125020)  
AUTHORS Holmes,S.E., Inglesoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.  
TITLE Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical  
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA  
JOURNAL  
FEATURES  
Source 1..125020  
location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16q24.3: between D16S520 and WI-12410"  
/note="Isolated from a patient with Huntington's  
Disease-like 2 (HD12)"  
complement(35581..35746)  
/rpt\_type=tandem  
/rpt\_unit=cig  
complement(<36507..>36887)  
/gene="JPH3"  
/note="synonym: JPH3"  
complement(<36507..>36887)  
/product="junctophillin 3"  
complement(<36507..36887)  
/gene="JPH3"

/note="component of the junctional complex between plasma  
membrane and endoplasmic reticulum"  
/codon\_start=1  
/product="junctophillin 3"  
/protein\_id="AA140941.1"  
/db\_xref="GI:17646245"  
/translation="MSSGRFPNDGGSYCGGMDKRAHGHVCTGPKGGEYTGWS  
HGFEYLTWPSGNTYGTGTAAGCRGIGESGKRYKEMTHGFRGIVREACG  
NGAYREGWSGLDDGIGETYSOG"  
BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others  
ORIGIN

Query Match 4.3%; Score 49; DB 9; Length 125020;  
Best Local Similarity 12.3%; Pred. No. 0.11;  
Matches 114; Conservative 385; Mismatches 419; Indels 10; Gaps 4;  
QY 147 GATTTCTGCGGGAAGCTGTGACCATGTCACCGTCCGAAGAGTTCTGATTCGAGAG 206  
Db 17869 GATCTCMAGKCMCTYKSKKKYASKGRKGTYYWRKSMARKMYRMAYSMR 17810  
QY 207 CCTTGCTTCGCGAAGGCTGATGCTCCCTGCTGTGCGGGAAGTCCGTCACA 266  
Db 17809 CAGMGAACGASRSWKCWGGYRKSAGSRSGTGYYWKKGGKGSMSKMKGSSTS 17750  
QY 267 TCGTCGACGAAATTTCTCAAGTGTGTCATCACTCGGAGCTGATCTGTCGAT 326  
Db 17749 RRRSSASCSYMW--GSMCMCSWMAKSYMYCYRMSSYSYYKCYSCWMSST 17692  
QY 327 GGTGACGATGACCATGCGCATACAGTGTGACGCGCTCCACGACTATGATGCTGAT 386  
Db 17691 SYSYSCWMSWMSGCMYKMKYSGMSSYSTSMYSSISCKKYCMSWYKCTMSKY 17632  
QY 387 TTGCGAGCTGGAAGCTGCGACCAACGCTGCTGTTGGCGTACGGAAGTGAATGTA 446  
Db 17631 RRRSSYVWGGKRAKKYKACGRRRMSWYKCAKMMWSYCWMSYKCTYKS 17572  
QY 447 CGGCTGCAACGATACGTGCGCGCTGTTGAATCACTACGGAATAAGCTGTGACGT 506  
Db 17571 STCYRGGYWSKTCYSAAGSRMYCMMSSSSSSWMSWMASSWCMGAGYRKR 17512  
QY 507 CCACGCTTCCGCGCTACTCATCATGATGATGATGATGATGATGATGATGATGATGAT 566  
Db 17511 RAAGMAGRRSKGRKSRGTMKRAAGSKTGSYSTSRSMKKKG----YKYSRGMKKK 17457  
QY 567 CTGCGATGCGATGTTGATGCGTGAATCTCAAACTGATATCTGCGGATTTGCTGCGCC 626  
Db 17456 KTCYCMWKKYKRYKRTSMCWYMYKSWGYRRCCKKKKCTGTGMSKSGYKSMRG 17397  
QY 627 ACTAGTAATGCGGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 686  
Db 17396 RGSYSTSCWKSCKWGSWYKMYKSKKRSMSGMSKGMRYAGRCYSSSWSTR 17337  
QY 687 CGCTGCTATCTGACGAATAATTCATCTGCGCTGTGATGATGATGATGATGATGATGATG 746  
Db 17336 KRRSKCYSYKSKYKGRKMGKMGKMSKYSMSKMSKYSKYSKGRKK 17277  
QY 747 AAAAAGTGTCATGATGCAACCTGCGACCTCGGAATTCCTGTTGATGACAGTGGCGG 806  
Db 17276 GMRSTKSKAKSMWAGSKCTYSSNNRRNNRMKGTGYMYRRARMNNGAAAA 17217  
QY 807 GGAATCACAAGGTTGGAACTGTGGG--AGAAATTTGATTTGATGATGATGATGATGATG 864  
Db 17216 GCTTCCCAATNGGGGAAAAAGGGGASRSRYSYKGRSSSBSYSGTGRRRMKSKCR 17157  
QY 865 GCAATGCGTGTCCCGATGATGCGCCGAGCTGATCTTGGGCAAGATCAAGAT 924  
Db 17156 GSRGKSMGNTRRSGSKTYSKAGSRGTYCMGWRKRRKCKMSKMK--MYSYRRRR 17098  
QY 925 GCTCAGATGTTGACGAGAAATCTGCCCCGATCATGATGATGATGATGATGATGATGATG 984  
Db 17097 WMTCKMCMYMWAMYCRMCMCMCKSCCGCYSMGSSYSYSGKSMGKSMYMRSSYS 17038  
QY 985 TCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1044

Qy	371	ACCATGAGCATGACCTTTCCGACAGCCCTGGAAAGCTGGACCAACAGCTGCTGTTGGGCTA	430
Db	974	CTGTTTGGAGTGGTGTGCTCTTCGCGGGGCGCGAGTCATCTTCGGGCGAAGGCGCTGAGCGT	915
Qy	431	CGGAACGTGAAATGTAACGCTCGACGCGAATACGTGCGCGCTGTGAAATCACTCACGAA	490
Db	914	CTBACACAGCTCCGTACACGAGGAGTGGAGGCGCTTCGTGCTGTGTCGAGTTACGACACGGGC	855
Qy	491	AAAACGGCTGGCAGCTCCACGTTACAGCGCTTACTCATGTTCAATGTTGACGTGAGTGA	548
Db	854	GCAATGGGTGGTCATGTGCATCTGCATGTGCTCGTGTTCCTGATGGCGGACTTCGGCGA	797
RESULT 14			
AF429315			
LOCUS	AF429315	125020 bp	DNA linear PRI 18-JAN-2002
DEFINITION	Homo sapiens junctophillin 3 (JPH3) gene, partial cds.		
ACCESSION	AF429315		
VERSION	AF429315.1	GI:17646244	
KEYWORDS			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1. (bases 1 to 125020)		
JOURNAL	Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,		
MEDLINE	Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,		
PUBMED	Potter,N.T., Ross,C.A. and Margolis,R.L.		
REFERENCE	A repeat expansion in the gene encoding junctophillin-3 is		
AUTHORS	associated with Huntington disease--like 2		
JOURNAL	Nat. Genet. 29 (4), 377-378 (2001)		
TITLE	2 (bases 1 to 125020)		
MEDLINE	Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.		
PUBMED	Direct Submission		
REFERENCE	Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical		
AUTHORS	Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1..125020		
SOURCE	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16q24.3; between D16S520 and WI-12410"		
	/note="Isolated from a patient with Huntington's		
	disease-like 2 (HDL2)"		
	complement (35581..35746)		
	/rpt_type=tandem		
	/rpt_unit=ctg		
	complement (<36507..>36887)		
	/gene="JPH3"		
	/note="Synonym: JP3"		
	complement (<36507..>36887)		
	/gene="JPH3"		
	/product="Junctophillin 3"		
	complement (<36507..36887)		
	/gene="JPH3"		
	/note="Component of the junctional complex between plasma		
	membrane and endoplasmic reticulum"		
	/codon_start=1		
	/product="Junctophillin 3"		
	/protein_id="P414094.1"		
	/db_xref="GI:17646245"		
	/translation="WSSGGRFNPDDGSGYCGWDEKAHGVCTPKGQGEYTGSMNS		
	HGEVGLVYTPSGNTYQGTMAQKRGILBKGKVVYGEWTHGKGYVRECAAG		
	NGAKYEGTWGSLQDGGTETPEDG"		
BASE COUNT	29056 a 32731 c 30696 g 28283 t 4254 others		
ORIGIN			
Query Match	4.1%;	Score 46.4;	DB 9; Length 125020;
Best Local Similarity	10.8%;	Pred. No. 0.52;	



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 12:01:45 ; Search time 281 Seconds  
(without alignments)  
10951.457 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1 atgacagcgtaagtcgtga.....taatacgttcgaagtga 1140

Sequence:

IDENTITY NUC

Scoring table: Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Database : N\_Geneseq\_19Jun03.\*

```
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1140	100.0	ABQ76122	Rhodococcus AN12 r
2	1140	100.0	ABQ76124	Rhodococcus AN12 d
3	1140	100.0	ABQ76126	Plasmid pRHR171 DN
4	1140	100.0	ABQ76125	Plasmid pRHR17 DN
5	51.6	4.5	AAV58945	B. breve essential
6	37.4	3.3	AAA00934	B. lactofermentum
7	37.4	3.3	AAA00951	B. lactofermentum
8	37.4	3.3	ABU49733	Brevibacterium lac

9	37.4	3.3	4447	24	ABU49734	Brevibacterium lac
10	37.4	3.3	4447	24	ABU22582	B. lactofermentum
11	37.4	3.3	4447	24	ABU22583	B. lactofermentum
12	34.8	3.1	1631	21	AA66016	E. coli proliferat
13	34.8	3.0	4590	22	AA424065	Yeast AOP9604-asso
14	34.6	3.0	1505	15	AA055750	Genomic clone G11F
15	34.6	3.0	1954	24	AB199602	Mouse ischaemic co
16	34.4	3.0	1173	22	AA426431	Pseudomonas sp hea
17	34.2	3.0	684	22	AA426442	T. versicolor pyrF
18	34.2	3.0	1314	24	ABQ90222	M. capsulatus gene
19	34.2	3.0	3448	22	AA26441	T. versicolor pyrF
20	34.2	3.0	44938	23	AA595525	Propionibacterium
21	34	3.0	534	24	ABQ41452	ABQ41452
22	34	3.0	534	24	ABQ41453	Oligonucleotide fo
23	33.6	2.9	786	25	ACA00393	C. glutamicum deri
24	33.6	2.9	832	24	ABQ20710	Oligonucleotide fo
25	33.6	2.9	832	24	ABQ20711	Oligonucleotide fo
26	33.6	2.9	1134	21	AA558791	DNA encoding the b
27	33.6	2.9	1134	24	AB553440	Degenerate DNA enc
28	33.6	2.9	1134	24	AA639776	C glutamicum codin
29	33.6	2.9	1338	22	AA66165	C glutamicum codin
30	33.6	2.9	349980	22	AA68527	DNA encoding novel
31	33.4	2.9	2730	23	AA887782	DNA encoding novel
32	33.4	2.9	3297	23	AA591437	DNA encoding novel
33	33.4	2.9	58985	25	AB259738	Human secreted pro
34	33.4	2.9	143601	25	AB22654	Human epidermal gr
35	33.2	2.9	549	21	AA05338	Cat flea head and
36	33.2	2.9	1780	19	AAV52605	Human oxidoreducta
37	33.2	2.9	1783	25	ACC46214	Human dltip intrac
38	33.2	2.9	1798	24	AB156822	Mouse interleukin-
39	33.2	2.9	1843	21	AA37046	Human PRO1265 (UNQ
40	33.2	2.9	1844	20	AA487257	CDNA clone encodin
41	33.2	2.9	1844	21	AA46906	CDNA encoding nove
42	33.2	2.9	1844	22	AA521481	Human CDNA sequenc
43	33.2	2.9	1844	22	AA029230	Human PRO1265 CDNA
44	33.2	2.9	1844	22	AA54260	DNA encoding prote
45	33.2	2.9	1844	22	AA091564	Human PRO1265 CDNA

#### ALIGNMENTS

RESULT 1	
ABQ76122	ABQ76122 standard; DNA; 1140 BP.
ID	
XX	ABQ76122;
AC	ABQ76122;
DT	13-JUN-2003 (first entry)
XX	
DE	Rhodococcus AN12 replication protein Rep DNA.
XX	
KW	Plasmid stability protein; replication protein; ethylene forming enzyme;
KW	carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
KW	polyhydroxyalkanoic acid synthase; PBA synthase; nitrile hydratase;
KW	alcohol dehydrogenase; terpene synthase; cholesterol oxidase; gene;
XX	shuttle vector; Rep; ds.
OS	Rhodococcus erythropolis.
XX	
PN	2002055709-A2.
XX	
PD	18-JUL-2002.
XX	
PF	12-DEC-2001; 2001WO-US47668.
XX	
PR	12-DEC-2000; 2000US-254868P.
XX	
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.
XX	
PI	Bismucci MG, Cheng Q, Kostichka KN, Tomb J;
XX	
DR	WPI; 2002-557827/59.



DR P-PSDB; ABB84278.  
 XX  
 PT New nucleic acid molecule encoding replication protein/plasmid  
 PT stability protein, useful in cloning and expression vectors,  
 PT particularly shuttle vectors for expression of heterologous genes in  
 PT Rhodococcus species  
 XX  
 PS Claim 2; Page 63-64; 96pp; English.  
 CC This invention describes a novel nucleic acid encoding a replication  
 CC protein or a plasmid stability protein. The product of the invention is  
 CC useful for expression of nucleic acid such as genes encoding enzymes  
 CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
 CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
 CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
 CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
 CC actinomyetales bacteria. The replication protein or plasmid stability  
 CC protein are useful in cloning and expression vectors and particularly in  
 CC shuttle vectors for the expression of homologous and heterologous genes  
 CC in Rhodococcus sp. This sequence encodes the Rhodococcus AN12 Rep  
 CC protein described in the disclosure of the invention.  
 CC  
 XX  
 SQ Sequence 1140 BP; 241 A; 287 C; 347 G; 265 T; 0 other;  
 Query Match 100.0%; Score 1140; DB 24; Length 1140;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGACCGAGTGAAGTGTGGAACACCTTTCCGGCAAAACCGGCTCCCGTCTGCTGCG 60  
 Db 1 ATGACCGAGTGAAGTGTGGAACACCTTTCCGGCAAAACCGGCTCCCGTCTGCTGCG 60  
 QY 61 TCCGATAGCGCGGCATCCGGACAGAACTGCGAACCCAACTTCAACAATCACCAGCTCA 120  
 Db 61 TCCGATAGCGCGGCATCCGGACAGAACTGCGAACCCAACTTCAACAATCACCAGCTCA 120  
 QY 121 GAAACATTTAACGCTGTGCGCCGCGCATTTCTGCGCTGGAACGGTGTGACCATTTGTCAAC 180  
 Db 121 GAAACATTTAACGCTGTGCGCCGCGCATTTCTGCGCTGGAACGGTGTGACCATTTGTCAAC 180  
 QY 181 GGTCCGAAAGGTTTGTGATTCGAGAGGCTTGTCTTCCGGAAGGGCTGTGATTCGCC 240  
 Db 181 GGTCCGAAAGGTTTGTGATTCGAGAGGCTTGTCTTCCGGAAGGGCTGTGATTCGCC 240  
 QY 241 TGCTGTGCGGGAAGGTCGATGCGACATGTCGACAGAAATTTCTCAAGTGTGTCTCAT 300  
 Db 241 TGCTGTGCGGGAAGGTCGATGCGACATGTCGACAGAAATTTCTCAAGTGTGTCTCAT 300  
 QY 301 CAACTCGGAGCTGATCTGTTGCGATGATGATGACATGCGCCATACAGCTGTGAC 360  
 Db 301 CAACTCGGAGCTGATCTGTTGCGATGATGATGACATGCGCCATACAGCTGTGAC 360  
 QY 361 CGGCTCCACGACCTATGAGCTGGAATTTCCGCGACCTGGAAGAGTGGACCAAGCTGTCT 420  
 Db 361 CGGCTCCACGACCTATGAGCTGGAATTTCCGCGACCTGGAAGAGTGGACCAAGCTGTCT 420  
 QY 421 CGTGTGCGTACGGAACGTGAATGTAAGCGCTGCGACGAGTACGAGTGTGAGTATC 480  
 Db 421 CGTGTGCGTACGGAACGTGAATGTAAGCGCTGCGACGAGTACGAGTGTGAGTATC 480  
 QY 481 ACTCAGGAAAAAAGCGCTGCGACGTCACGTCACGTCGACTCATGTTCAAGTGTGAC 540  
 Db 481 ACTCAGGAAAAAAGCGCTGCGACGTCACGTCACGTCGACTCATGTTCAAGTGTGAC 540  
 QY 541 GTGAGTGAAGACATCTCGAATCTTTCGAGTCGATGTTGATCGGTGAGCTTCCAAA 600  
 Db 541 GTGAGTGAAGACATCTCGAATCTTTCGAGTCGATGTTGATCGGTGAGCTTCCAAA 600  
 QY 601 CTGCTATCTCTGGAGATTGCTGCGCACTAGTAATTCGGTGTGCTCGATGATGAGAA 660  
 Db 601 CTGCTATCTCTGGAGATTGCTGCGCACTAGTAATTCGGTGTGCTCGATGATGAGAA 660  
 QY 661 ATCGCGGTGAAGCTGATCAAGTTCGCTGCGATCTGACGAAATTTGACCTGCGGTT 720

Db 661 ATCGCGGTGAAGCTGATCAAGTTCGCTGCGATCTGACGAAATTTGACCTGCGGTT 720  
 QY 721 GGTATGAGAGTGTGATGAGCGCAAGAAAGGTCGATGAGCAACCGTCAACCTCG 780  
 Db 721 GGTATGAGAGTGTGATGAGCGCAAGAAAGGTCGATGAGCAACCGTCAACCTCG 780  
 QY 781 GAAATCGCTGTTGATGATGAGTGGAGCGGATCCCAAGCGTTGGAATCTGAGCAATTT 840  
 Db 781 GAAATCGCTGTTGATGATGAGTGGAGCGGATCCCAAGCGTTGGAATCTGAGCAATTT 840  
 QY 841 GAGTTTGTTCGATGAGGAGCTGCGCAATCGCGTGTCCGTCGATGAGTGGCGAGCT 900  
 Db 841 GAGTTTGTTCGATGAGGAGCTGCGCAATCGCGTGTCCGTCGATGAGTGGCGAGCT 900  
 QY 901 GGTCTTGGGGCGAATAACAGATGCTCAGATCGTGAAGAGAAATTCGCCCGCTC 960  
 Db 901 GGTCTTGGGGCGAATAACAGATGCTCAGATCGTGAAGAGAAATTCGCCCGCTC 960  
 QY 961 ATGTTTGCATCATTCCTCGGCGGATGATGATGATTCGACCTTGTGCGCTTAAGTC 1020  
 Db 961 ATGTTTGCATCATTCCTCGGCGGATGATGATGATTCGACCTTGTGCGCTTAAGTC 1020  
 QY 1021 TTGCGCGAGATCTCTCGAATCTGTCGAGCTGCGCGCACTTGGAAAAATCTTGTGATCAC 1080  
 Db 1021 TTGCGCGAGATCTCTCGAATCTGTCGAGCTGCGCGCACTTGGAAAAATCTTGTGATCAC 1080  
 QY 1081 TTGCTATTCGATTCGCGGAGCGATGTCGCGGCGCCCGTAATATTCGTTTGGCAAGTGA 1140  
 Db 1081 TTGCTATTCGATTCGCGGAGCGATGTCGCGGCGCCCGTAATATTCGTTTGGCAAGTGA 1140  
 RESULT 2  
 ABQ76124/c  
 ID ABQ76124 standard; DNA; 6334 BP.  
 XX  
 AC ABQ76124;  
 AC  
 PT 13-JAN-2003 (first entry)  
 XX  
 DE Rhodococcus AN12 derived plasmid pAN12 DNA.  
 XX  
 KW Plasmid stability protein; replication protein; ethylene forming enzyme;  
 KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
 KW polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;  
 KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;  
 KW shuttle vector; circular; ds.  
 XX  
 OS Rhodococcus erythropolis.  
 OS Synthetic.  
 OS  
 PN WO200255709-A2.  
 PN  
 PD 18-JUL-2002.  
 PD  
 XX 12-DEC-2001; 2001WO-US47868.  
 PF  
 PR 12-DEC-2000; 2000US-254868P.  
 PR  
 XX (DUPO ) DU PONT DE NEMOURS & CO E. I.  
 PA  
 XX Bramucci MG, Cheng Q, Kostichka KN, Tomb J;  
 PI  
 DR WPI; 2002-557827/59.  
 XX  
 XX New nucleic acid molecule encoding replication protein/plasmid  
 PT stability protein, useful in cloning and expression vectors,  
 PT particularly shuttle vectors for expression of heterologous genes in  
 PT Rhodococcus species  
 PS Claim 19; Page 68-71; 96pp; English.  
 PS  
 CC This invention describes a novel nucleic acid encoding a replication

CC protein or a plasmid stability protein. The product of the invention is  
CC useful for expression of nucleic acid such as genes encoding enzymes  
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
CC Actinomycetales bacteria. The replication protein or plasmid stability  
CC protein are useful in cloning and expression vectors and particularly in  
CC shuttle vectors for the expression of homologous and heterologous genes  
CC in Rhodococcus sp. This sequence represents the Rhodococcus AN12 derived  
CC plasmid PAN12 DNA described in the disclosure of the invention.  
XX  
SQ Sequence 6334 BP; 1386 A; 1807 C; 1681 G; 1460 T; 0 other;

Query Match 100.0%; Score 1140; DB 24; Length 6334;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCGAGCTAAGTGTGTAACACCTTTCGGGAAAGACCGGCTCCCGTCTCGTGTG 60  
DB ATACACGAGCTAAGTGTGTAACACCTTTCGGGAAAGACCGGCTCCCGTCTCGTGTG 2992  
QY 61 TCCGATTAAGCGCGGCAATCCGGGACGAACTGCGAACCCTTCAACCAATCACCGTCA 120  
DB TCCGATTAAGCGCGGCAATCCGGGACGAACTGCGAACCCTTCAACCAATCACCGTCA 2932  
QY 121 GAAACATTTAACCGCTGTGCGCGGCGGATTTCTGCGGTGAACGCTGTGCAAC 180  
DB GAAACATTTAACCGCTGTGCGCGGCGGATTTCTGCGGTGAACGCTGTGCAAC 2932  
QY 2931 GAAACATTTAACCGCTGTGCGCGGCGGATTTCTGCGGTGAACGCTGTGCAAC 2872  
QY 181 GGTCCGAAAGCTTGTGATTCGAGAGGCTTCTGCGGTGAACGCTGTGCAAC 240  
DB GGTCCGAAAGCTTGTGATTCGAGAGGCTTCTGCGGTGAACGCTGTGCAAC 2872  
QY 2871 GGTCCGAAAGCTTGTGATTCGAGAGGCTTCTGCGGTGAACGCTGTGCAAC 2812  
QY 241 TGCTGTGCGGAAAGCTGCGTGAACATCGTGAACGAAATTTCTCAAGTTGTGTCAT 300  
DB TGCTGTGCGGAAAGCTGCGTGAACATCGTGAACGAAATTTCTCAAGTTGTGTCAT 2752  
QY 2811 TGCTGTGCGGAAAGCTGCGTGAACATCGTGAACGAAATTTCTCAAGTTGTGTCAT 2752  
QY 301 CAACTCGGAGCTGATCTGTGCGATGTGACATGACATGCGCCATACAGCTGTGAC 360  
DB CAACTCGGAGCTGATCTGTGCGATGTGACATGACATGCGCCATACAGCTGTGAC 2751  
QY 2751 CAACTCGGAGCTGATCTGTGCGATGTGACATGACATGCGCCATACAGCTGTGAC 2692  
QY 361 CGGCTCCAGCACTATGATGATGATCTTTCGCGAGCTTGAAGTGTGCAACAGTGT 420  
DB CGGCTCCAGCACTATGATGATGATCTTTCGCGAGCTTGAAGTGTGCAACAGTGT 2691  
QY 2691 CGGCTCCAGCACTATGATGATGATCTTTCGCGAGCTTGAAGTGTGCAACAGTGT 2632  
QY 421 CGTTCGCTGACGAAAGCTGAAATCTAGCGTGAACGATGACGCGCTGTGTAATC 480  
DB CGTTCGCTGACGAAAGCTGAAATCTAGCGTGAACGATGACGCGCTGTGTAATC 2631  
QY 2631 CGTTCGCTGACGAAAGCTGAAATCTAGCGTGAACGATGACGCGCTGTGTAATC 2572  
QY 481 ACTCAGGAAAGAAAGCGCTGAGCACTGCAAGTTCACGCGTACATGTCAGTGTGAC 540  
DB ACTCAGGAAAGAAAGCGCTGAGCACTGCAAGTTCACGCGTACATGTCAGTGTGAC 2571  
QY 2571 ACTCAGGAAAGAAAGCGCTGAGCACTGCAAGTTCACGCGTACATGTCAGTGTGAC 2512  
QY 541 GTGATGAGAAACATCTCTCGAATCTTTCGATGCGATGCGATGCGTGTGCAAA 600  
DB GTGATGAGAAACATCTCTCGAATCTTTCGATGCGATGCGATGCGTGTGCAAA 2511  
QY 2511 GTGATGAGAAACATCTCTCGAATCTTTCGATGCGATGCGATGCGTGTGCAAA 2452  
QY 601 CTGCTATCTCTGGGATTTGCTGCGGCACTAGTAATTCGATGCGATGCGTGTGCAAA 660  
DB CTGCTATCTCTGGGATTTGCTGCGGCACTAGTAATTCGATGCGATGCGTGTGCAAA 2451  
QY 2451 CTGCTATCTCTGGGATTTGCTGCGGCACTAGTAATTCGATGCGATGCGTGTGCAAA 2392  
QY 661 ATCGCGGAGTGAAGCTGATCAAGTTCGCTGCGTATGCAAGAAATTTGATCGGCTT 720  
DB ATCGCGGAGTGAAGCTGATCAAGTTCGCTGCGTATGCAAGAAATTTGATCGGCTT 2391  
QY 2391 ATCGCGGAGTGAAGCTGATCAAGTTCGCTGCGTATGCAAGAAATTTGATCGGCTT 2332  
QY 721 GGTATGAGAGTGTGTAGTGGCGACGAAAGTGTGTGACATGCGAACCGTGTGAC 780  
DB GGTATGAGAGTGTGTAGTGGCGACGAAAGTGTGTGACATGCGAACCGTGTGAC 2331  
QY 2331 GGTATGAGAGTGTGTAGTGGCGACGAAAGTGTGTGACATGCGAACCGTGTGAC 2272  
QY 781 GAAATGCTGTTTATGATGAGTGGCGGAGATCCAAACGTTTGAACGTTGCGCGAATTT 840  
DB GAAATGCTGTTTATGATGAGTGGCGGAGATCCAAACGTTTGAACGTTGCGCGAATTT 2271

QY 841 GAGTTTGTTCATGAGAGCTGCGGCAATCGGCTGTCCCTGATTTGCTGCCGAGCT 900  
DB GAGTTTGTTCATGAGAGCTGCGGCAATCGGCTGTCCCTGATTTGCTGCCGAGCT 2152  
QY 901 GGTCTGGGGGAGAACTAACAGATGCTGATGCTTGAAGCAGAAAGAACTGCCCGGTC 960  
DB GGTCTGGGGGAGAACTAACAGATGCTGATGCTTGAAGCAGAAAGAACTGCCCGGTC 2151  
QY 961 ATGTTTGCATGATTCGCGCGGAGATCGTGAATGATGATTCGACCTTGTGCGCTTAC 1020  
DB ATGTTTGCATGATTCGCGCGGAGATCGTGAATGATGATTCGACCTTGTGCGCTTAC 2091  
QY 1021 TTGCGGAGATCTCTGAGATCTGCGAAGCTGCGCGCACTTTGGGAAATCTTGTGAT 1080  
DB TTGCGGAGATCTCTGAGATCTGCGAAGCTGCGCGCACTTTGGGAAATCTTGTGAT 2031  
QY 2031 TTGCGGAGATCTCTGAGATCTGCGAAGCTGCGCGCACTTTGTGATGATC 1972  
QY 1081 TTGCAATTATGATGCTGCGCGAGCGATGCGCGGCGCGGATTAATCGTGTGCAAGTA 1140  
DB TTGCAATTATGATGCTGCGCGAGCGATGCGCGGCGCGGATTAATCGTGTGCAAGTA 1971

## RESULT 3

ABQ76126/c  
ID ABQ76126 standard; DNA; 9652 BP.

ABQ76126;

13-JAN-2003 (first entry)

Plasmid pRHR171 DNA.

XX Plasmid stability protein; replication protein; ethylene forming enzyme;  
XX carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
XX polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;  
XX alcohol dehydrogenase; terpene synthase; cholesterol oxidase;  
XX shuttle vector; circular; ds.

XX Synthetic.

XX NO200255709-A2.

XX 18-JUL-2002.

XX 12-DEC-2001; 2001WO-US47868.

XX 12-DEC-2000; 2000US-254868P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Bramucci MG, Cheng Q, Kostichka KN, Tomb J;

XX WPI; 2002-557827/59.

XX New nucleic acid molecule encoding replication protein/plasmid  
XX stability protein, useful in cloning and expression vectors,  
XX particularly shuttle vectors for expression of heterologous genes in  
XX Rhodococcus species -

XX Claim 27; Page 72; 96pp; English.

XX This invention describes a novel nucleic acid encoding a replication  
XX protein or a plasmid stability protein. The product of the invention is  
XX useful for expression of nucleic acid such as genes encoding enzymes  
XX involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
XX acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
XX hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
XX dehydrogenase, terpene synthases, and cholesterol oxidase in an  
XX Actinomycetales bacteria. The replication protein or plasmid stability  
XX protein are useful in cloning and expression vectors and particularly in  
XX shuttle vectors for the expression of homologous and heterologous genes  
XX in Rhodococcus sp. This sequence represents the Plasmid pRHR171 DNA  
XX described in the disclosure of the invention.

XX Sequence 9652 BP; 2097 A; 2729 C; 2559 G; 2267 T; 0 other;  
SQ  
Query Match 100.0%; Score 1140; DB 24; Length 9652;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACGAGCGTAAGTGTGTAACACCTTTCCGCGAAAGACCGGCTTCGTCCTGCTGTCG 60  
DB 5052 ATGACGAGCGTAAGTGTGTAACACCTTTCCGCGAAAGACCGGCTTCGTCCTGCTGTCG 4993  
QY 61 TCCGATTAAGCGCGGCAATCCGCGACGAACTGCGACCCAACTTAAACAATCAACAGTCA 120  
DB 4992 TCCGATTAAGCGCGCAATCCGCGACGAACTGCGACCCAACTTAAACAATCAACAGTCA 4933  
QY 121 GAACATTTAAGCGCTGTGGCCGCGCATTTCTGCGGTGAACGGTGTGACATTTGTCAAC 180  
DB 4932 GAACATTTAAGCGCTGTGGCCGCGCATTTCTGCGGTGAACGGTGTGACATTTGTCAAC 4873  
QY 181 GGTCCGAAAGGTTCTGGAATTCGAGGACCTTCGTTCTGCGGAAAGGGCTGGAATCTGCCCC 240  
DB 4872 GGTCCGAAAGGTTCTGGAATTCGAGGACCTTCGTTCTGCGGAAAGGGCTGGAATCTGCCCC 4813  
QY 241 TGTCTGTGCGGAAAAAGTCCGTGCAATCGTGTGCAAGAAATTTCTCAAGTTGTGCTCAT 300  
DB 4812 TGTCTGTGCGGAAAAAGTCCGTGCAATCGTGTGCAAGAAATTTCTCAAGTTGTGCTCAT 4753  
QY 301 CAACCTGGAGCGATGATCTGTGCGATGTGATGATGATGATGATGATGATGATGATGATGAT 360  
DB 4752 CAACCTGGAGCGATGATCTGTGCGATGTGATGATGATGATGATGATGATGATGATGATGAT 4693  
QY 361 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
DB 4692 CGGCTCCACGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4633  
QY 421 CGTTGGCGTTCGGAACGTGAAATGTATGAGCTGTGCAAGCAATGCGTGTGTTGAATTC 480  
DB 4632 CGTTGGCGTTCGGAACGTGAAATGTATGAGCTGTGCAAGCAATGCGTGTGTTGAATTC 4573  
QY 481 ACTCAAGGAAAAAGCGCTGTGCAAGTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 540  
DB 4572 ACTCAAGGAAAAAGCGCTGTGCAAGTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 4513  
QY 541 GTGATGAGAAACATCTCGAATCTTCTCGAATGCGATGTTGATCGGTGACTTCCAAA 600  
DB 4512 GTGATGAGAAACATCTCGAATCTTCTCGAATGCGATGTTGATCGGTGACTTCCAAA 4453  
QY 601 CTGATATCTCTGGGATTTGCTGCGGCACTACGTAATTCGGGTGTCTCGATGTACGAAAG 660  
DB 4452 CTGATATCTCTGGGATTTGCTGCGGCACTACGTAATTCGGGTGTCTCGATGTACGAAAG 4393  
QY 661 ATGCGGCGTGAAGCGTGAATCAAGTTCTGCGTGGCGTATCGAAGAAATTTGATCGGCGTT 720  
DB 4392 ATGCGGCGTGAAGCGTGAATCAAGTTCTGCGTGGCGTATCGAAGAAATTTGATCGGCGTT 4333  
QY 721 GGTATGAGAGTTGTGATGTGCGACGAAAAAGTGTGCAATGTGCAACCGTGCACCTGCG 780  
DB 4332 GGTATGAGAGTTGTGATGTGCGACGAAAAAGTGTGCAATGTGCAACCGTGCACCTGCG 4273  
QY 781 GAAATCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
DB 4272 GAAATCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4213  
QY 841 GAGTTTGGTTTGAAGGAGCGTGGGCAATCGCGTGTGATCGCGTGTGATCGCGGAGCT 900  
DB 4212 GAGTTTGGTTTGAAGGAGCGTGGGCAATCGCGTGTGATCGCGTGTGATCGCGGAGCT 4153  
QY 901 GGTCTTGGGCGGAGAACTAAACAGATGCTCAATCGTTGACAGAAAGATCTGCCCCGCTC 960  
DB 4152 GGTCTTGGGCGGAGAACTAAACAGATGCTCAATCGTTGACAGAAAGATCTGCCCCGCTC 4093  
QY 961 ATGTTTGGATCATCTCCGCGCGGATCGTGTGATGATGATGATGATGATGATGATGATGATGAT 1020

DB 4092 ATGTTTGGATCATCTCCGCGCGGATCGTGTGATGATGATGATGATGATGATGATGATGATGAT 4033  
QY 1021 TTCCGCGAGATCTTCGACCTGTGCAAGCTGCGGCACTTGGAAAAATCTTGATGATCA 1080  
DB 4032 TTCCGCGAGATCTTCGACCTGTGCAAGCTGCGGCACTTGGAAAAATCTTGATGATCA 3973  
QY 1081 TTGCAATTAATGATTCCTCCGCGAGCGGATGTGCGGCCCCCGAATTAATTCGTTTCCGAAAGTA 1140  
DB 3972 TTGCAATTAATGATTCCTCCGCGAGCGGATGTGCGGCCCCCGAATTAATTCGTTTCCGAAAGTA 3913

## RESULT 4

AB076125/c  
ID AB076125 standard; DNA; 11241 BP.

AB076125;

13-JAN-2003 (first entry)

Plasmid pRHB17 DNA.

Plasmid stability protein; replication protein; ethylene forming enzyme;

carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;

polyhydroxyalkanoic acid synthase; pHA synthase; nitrile hydratase;

alcohol dehydrogenase; terpene synthase; cholesterol oxidase;

shuttle vector; circular; ds.

Synthetic.

200255709-A2.

18-JUL-2002.

12-DEC-2001; 2001MO-US47868.

12-DEC-2000; 2000US-254868P.

(DUPO) DU POINT DE MEMOIRS & CO E. I.

Bramucci MG, Cheng Q, Kostichka KM, Tomb J;

WPI; 2002-557827/59.

Claim 26; Page 71-72; 96pp; English.

This invention describes a novel nucleic acid encoding a replication

protein or a plasmid stability protein. The product of the invention is

useful for expression of nucleic acid such as genes encoding enzymes

involved in the production of isoprenoid molecules, polyhydroxyalkanoic

acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile

hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol

dehydrogenase, terpene synthases, and cholesterol oxidase in an

Actinomycetales bacteria. The replication protein or plasmid stability

protein are useful in cloning and expression vectors and particularly in

shuttle vectors for the expression of homologous and heterologous genes

in Rhodococcus sp. This sequence represents the Plasmid pRHB17 DNA

described in the disclosure of the invention.

SQ Sequence 11241 BP; 2517 A; 3109 C; 2945 G; 2670 T; 0 other;

Query Match 100.0%; Score 1140; DB 24; Length 11241;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCGTAAGTGTGTAACACCTTTCCGCGAAAGACCGGCTTCGTCCTGCTGTCG 60  
DB 6641 ATGACGAGCGTAAGTGTGTAACACCTTTCCGCGAAAGACCGGCTTCGTCCTGCTGTCG 6582

```

QY 61 TCCGATTAAGCGGGCATCCGGGCGAAGCTGCGAACCCTTCAACAAATCACCACTCA 120
DB 6581 TCCGATTAAGCGGGCATCCGGGCGAAGCTGCGAACCCTTCAACAAATCACCACTCA 6522
QY 121 GAAACATTTAAAGCCCTGTCGGCCCGGATTTCTGCGGTGAACGCTGTGACATTTGCAAC 180
DB 6521 GAAACATTTAAAGCCCTGTCGGCCCGGATTTCTGCGGTGAACGCTGTGACATTTGCAAC 6462
QY 181 GGTCCGAAAAGTTCTGGAATTCGAGGCTTCGTTCTGCGGAAAAGGCTGATCTCTCCCC 240
DB 6461 GGTCCGAAAAGTTCTGGAATTCGAGGCTTCGTTCTGCGGAAAAGGCTGATCTCTCCCC 6402
QY 241 TGGTGTGCGGAAAAGTCCGTCGACATCGTCGACAGCAAAATTTCTCAAGTTGTGCTCAT 300
DB 6401 TGGTGTGCGGAAAAGTCCGTCGACATCGTCGACAGCAAAATTTCTCAAGTTGTGCTCAT 6342
QY 301 CAACCTGGGACTGTGATCTGTGCGATGTGACATGACATGACATGACATGACATGACATGAC 360
DB 6341 CAACCTGGGACTGTGATCTGTGCGATGTGACATGACATGACATGACATGACATGACATGAC 6282
QY 361 CGGCTCCAGCACTATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 420
DB 6281 CGGCTCCAGCACTATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 6222
QY 421 CGTTGGCGTACCGGAACTGTAAGTATGACGCTGCGACGATACGTCGCGCTGTGAAATC 480
DB 6221 CGTTGGCGTACCGGAACTGTAAGTATGACGCTGCGACGATACGTCGCGCTGTGAAATC 6162
QY 481 ACTCAGGAAAAAAGCGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCTGCG 540
DB 6161 ACTCAGGAAAAAAGCGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCTGCGACGCTGCG 6102
QY 541 GTGAGTGAAGAACATCTCGATCTCTCGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 6101 GTGAGTGAAGAACATCTCGATCTCTCGATGATGATGATGATGATGATGATGATGATGATGAT 6042
QY 601 CTGGAATCTCTGGAATTTGTCGCGCACTACGTAATTCGAGTGTCTGATGATGATGATGATG 660
DB 6041 CTGGAATCTCTGGAATTTGTCGCGCACTACGTAATTCGAGTGTCTGATGATGATGATGATG 5982
QY 661 ATGCGGCGTGAAGCTGATCAAGTTCTGCGCTGCTGATCTGACGAAAATTCGATCTGCGCTT 720
DB 5981 ATGCGGCGTGAAGCTGATCAAGTTCTGCGCTGCTGATCTGACGAAAATTCGATCTGCGCTT 5922
QY 721 GGTATGAGAGTTGTTGTTGTCGCGACGCGAAAAAGTGTGCAATGCGCAACCGTGCACCTTG 780
DB 5921 GGTATGAGAGTTGTTGTTGTCGCGACGCGAAAAAGTGTGCAATGCGCAACCGTGCACCTTG 5862
QY 781 GAAATGCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 5861 GAAATGCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5802
QY 841 GAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 5801 GAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5742
QY 901 GGTCTTGGGCGCAACATCAAGATGCTCAGATGCTGAGATGCTGAGATGCTGAGATGCTGAGAT 960
DB 5741 GGTCTTGGGCGCAACATCAAGATGCTCAGATGCTGAGATGCTGAGATGCTGAGATGCTGAGAT 5682
QY 961 ATGTTGCGATTCATTCGCGCGCATCGTGATGATGATGATGATGATGATGATGATGATGATG 1020
DB 5681 ATGTTGCGATTCATTCGCGCGCATCGTGATGATGATGATGATGATGATGATGATGATGATG 5622
QY 1021 TTGCGCGAGATTCCTCGGACTGTGCGAAGCTGCGCGCATTTGGGAAAAATCTTGTGATCAC 1080
DB 5621 TTGCGCGAGATTCCTCGGACTGTGCGAAGCTGCGCGCATTTGGGAAAAATCTTGTGATCAC 5562
QY 1081 TTGCAATTATGATTTGCCCGGAGGAGATGCGCGCCCGGCAATAATGCGTTGCCAGATGA 1140
DB 5561 TTGCAATTATGATTTGCCCGGAGGAGATGCGCGCCCGGCAATAATGCGTTGCCAGATGA 5502

```

```

RESULT 5
AAV58945
ID AAV58945 standard; DNA; 2297 BP.
XX
AC AAV58945;
XX
DT 05-JAN-1999 (first entry)
XX
DE B. breve essential region gene.
XX
ES Essential region gene; shuttle vector; ds.
XX
OS Bifidobacterium breve.
XX
FH Key 387..392 Location/Qualifiers
FT -35_signal /tag= a
FT -10_signal 410..415
FT RBS /tag= b
FT CDS 525..530
FT CDS 539..2023
FT CDS /tag= d
XX
PN JP10262670-A.
XX
PD 06-OCT-1998.
XX
PF 27-MAR-1997; 97JP-0091387.
XX
PR 27-MAR-1997; 97JP-0091387.
XX
PA (HONS ) YAKULT HONSHA KK.
XX
DR MPI: 1998-587288/50.
DR P-PSDB; AAW73071.
XX
PT Shuttle vector for a Bifidobacterium species - contains specific
PE elements from existing plasmids e.g. pNB1 of ATCC 15698
XX
PS Disclosure; Page 14-15; 17pp; Japanese.
XX
CC This sequence represents the essential region gene from plasmid pNB1
CC used in the vector of the invention. The vector is a shuttle vector for a
CC Bifidobacterium, and contains: (a) a replicated essential region
CC originated from a plasmid pNB1 of Bifidobacterium breve ATCC 15698;
CC (b) a replicated essential region originated from a plasmid of E. coli;
CC and (c) an antibiotic-resistant gene originating from a plasmid of E. coli
CC and an antibiotic-resistant gene functioning by a Bifidobacterium. The
CC shuttle vector can be used to transform various Bifidobacteria species.
XX
SQ Sequence 2297 BP; 524 A; 659 C; 689 G; 425 T; 0 other;
Query Match 4.5%; Score 51.6; DB 19; Length 2297;
Best Local Similarity 52.9%; Pred. No. 1.8e-05;
Matches 111; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 394 GCTGGAAGCTGCGCAACGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 453
DB 1184 GCTGGAAGCTGCGCAACGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1243
QY 454 GACGATACGTGCGCTGTTGAAATCACTCAGGAAAAAGCGCTGCGACGTCAGCTT 513
DB 1244 AGGGGTTTGTCCGCCGATGTAATCAGTCGATGTAAGCGCTGCGACCTTCATTT 1303
QY 514 CACGCGTACTCATGTTCAAGTGTGACGTAAGTGAAGACATCTCGAATCTTCTCGGAT 573
DB 1304 CATTCGATGATGTTCTCGATGCGCATGCGATGCGAGATGCGAGGCAATGACAGCAA 1363
QY 574 GCGATGTTGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 603
DB 1364 TGGCTGCTGATGCTGGAAGAACATGCTC 1393

```

```
RESULT 6
ID AAA90934
AC AAA90934 standard; DNA; 4447 BP.
XX
AC AAA90934;
XX
DT 15-JAN-2001 (first entry)
XX
DE B. lactofermentum p48K coding sequence.
XX
KM Temperature sensitive plasmid; TSRCR; protein production;
XX KM Temperature sensitive replication control region; p48K; ds.
XX OS Brevibacterium lactofermentum.
XX FH Key Location/Qualifiers
XX FT CDS 1318..2601
XX FT /*tag= a
XX
XX EP1038966-A1.
XX
XX PD 27-SEP-2000.
XX
XX PF 16-MAR-2000; 2000EP-0105326.
XX
XX PR 16-MAR-1999; 99JP-00689896.
XX
XX PA (AJIN ) AJINOMOTO CO INC.
XX
XX PI Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX
XX DR WPI; 2000-573832/54.
XX
XX DR P-PSDB; AAY97536.
XX
XX PT Plasmids containing a temperature sensitive replication control regions
XX PT useful for breeding microorganisms for the production of amino acids by
XX PT fermentation -
XX
XX PS Claim 7; Page 14-17; 29pp; English.
XX
XX CC This sequence encodes the Brevibacterium lactofermentum p48K protein.
XX CC The invention relates to a plasmid containing a temperature sensitive
XX CC replication control region (TSRCR) and a marker gene (MG). The TSRCR is
XX CC derived from plasmid pAM330 harboured by Brevibacterium lactofermentum
XX CC (ATCC 13869) and allows the plasmid to replicate autonomously at a low
XX CC temperature but does not allow the plasmid to replicate autonomously at
XX CC an elevated temperature in coryneform bacteria within a temperature
XX CC range in which the bacteria can grow. The plasmid can be used for
XX CC modifying a chromosomal gene in a coryneform bacterium, which may be used
XX CC for the production of useful substances, such as amino acids, by
XX CC fermentation, to change their genetic traits. Therefore, the plasmid can
XX CC be used for breeding microorganisms for the production of amino acids by
XX CC fermentation. The plasmid comprises a TSRCR that allows the plasmid to
XX CC replicate autonomously at an elevated temperature in coryneform
XX CC bacteria within a temperature range in which the bacteria can grow.
XX
XX SQ Sequence 4447 BP; 824 A; 997 C; 1378 G; 1248 T; 0 other;
XX
XX Query Match 3.3%; Score 37.4; DB 21; Length 4447;
XX Best Local Similarity 53.0%; Pred. No. 0.87;
XX Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
XX
XX DB 493 AACGGCTGGACGCTCACGCTTACGCGCTACTCATGTTGAGTGAGCGTAGAAGC 552
XX |||||
XX DB 1795 AACGGTTGGACATTGACCGCAACATGCTGTTTCTTGATGCTCACTGTGACGAT 1854
XX |||||
XX QY 553 ATCTCGAATCTTCTCGATGAGATGTTGATCGGTGACCTTCCAACTCGTATCTTG 612
XX |||||
XX DB 1855 GAACCTAAGGCGTTGAGGATTCATGTTTCCCGCTGCTGCTGATGTTAAGGCC 1914
XX |||||
XX QY 613 GGATTGCTGGCGCACTAGCTATTGGGGTG 643
XX |||||
```

```
DB 1915 GGTATGACGCGCCATGCTGTGACGACGGGG 1945
XX
XX RESULT 7
XX ID AAA90951
XX AC AAA90951 standard; DNA; 4447 BP.
XX
XX AC AAA90951;
XX
XX DT 15-JAN-2001 (first entry)
XX
XX DE B. lactofermentum pSF6 coding sequence.
XX
XX KM Temperature sensitive plasmid; TSRCR; protein production;
XX KM Temperature sensitive replication control region; pSF6; ds.
XX OS Brevibacterium lactofermentum.
XX FH Key Location/Qualifiers
XX FT CDS 1318..2601
XX FT /*tag= a
XX
XX EP1038966-A1.
XX
XX PD 27-SEP-2000.
XX
XX PF 16-MAR-2000; 2000EP-0105326.
XX
XX PR 16-MAR-1999; 99JP-00689896.
XX
XX PA (AJIN ) AJINOMOTO CO INC.
XX
XX PI Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX
XX DR WPI; 2000-573832/54.
XX
XX DR P-PSDB; AAY97537.
XX
XX PT Plasmids containing a temperature sensitive replication control regions
XX PT useful for breeding microorganisms for the production of amino acids by
XX PT fermentation -
XX
XX PS Disclosure; Page 19-22; 29pp; English.
XX
XX CC This sequence encodes the Brevibacterium lactofermentum pSF6 protein.
XX CC The invention relates to a plasmid containing a temperature sensitive
XX CC replication control region (TSRCR) and a marker gene (MG). The TSRCR is
XX CC derived from plasmid pAM330 harboured by Brevibacterium lactofermentum
XX CC (ATCC 13869) and allows the plasmid to replicate autonomously at a low
XX CC temperature but does not allow the plasmid to replicate autonomously at
XX CC an elevated temperature in coryneform bacteria within a temperature
XX CC range in which the bacteria can grow. The plasmid can be used for
XX CC modifying a chromosomal gene in a coryneform bacterium, which may be used
XX CC for the production of useful substances, such as amino acids, by
XX CC fermentation, to change their genetic traits. Therefore, the plasmid can
XX CC be used for breeding microorganisms for the production of amino acids by
XX CC fermentation. The plasmid comprises a TSRCR that allows the plasmid to
XX CC replicate autonomously at an elevated temperature in coryneform
XX CC bacteria within a temperature range in which the bacteria can grow.
XX
XX SQ Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;
XX
XX Query Match 3.3%; Score 37.4; DB 21; Length 4447;
XX Best Local Similarity 53.0%; Pred. No. 0.87;
XX Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
XX
XX DB 493 AACGGCTGGACGCTCACGCTTACGCGCTACTCATGTTGAGTGAGCGTAGAAGC 552
XX |||||
XX DB 1795 AACGGTTGGACATTGACCGCAACATGCTGTTTCTTGATGCTCACTGTGACGAT 1854
XX |||||
XX QY 553 ATCTCGAATCTTCTCGATGAGATGTTGATCGGTGACCTTCCAACTCGTATCTTG 612
XX |||||
XX DB 1855 GAACCTAAGGCGTTGAGGATTCATGTTTCCCGCTGCTGCTGATGTTAAGGCC 1914
XX |||||
```

QY 613 GGATTTCGCCCACTAGTATTCGGG 643  
DB 1915 GGATGACGCGCCACTGCGTGACGCGG 1945

## RESULT 8

ABL49733

ID ABL49733 standard; DNA; 4447 BP.

AC ABL49733;

DT 29-MAY-2002 (first entry)

DE Brevibacterium lactofermentum pSFK6 nucleotide sequence SEQ ID NO:5.

KM Kanamycin resistant gene; mutation; high temperature resistance;

KM coryneform microbe; transformation; gene; ds.

OS Brevibacterium lactofermentum.

FH Key Location/Qualifiers

FT CDS 1318..2601

FT /\*tag= a

FT /product= "pSFK6"

PN JP2002017362-A.

PD 22-JAN-2002.

PF 05-JUL-2000; 2000JP-0204236.

PR 05-JUL-2000; 2000JP-0204236.

PA (AJIN ) AJINOMOTO KK.

DR MPI; 2002-263243/31.

DR P-PSDB; ABB06341.

PT Transformation of high temperature-resistant coryneform microbe and

PT transformant -

PS Example; Page 16-18; 28pp; Japanese.

CC The present invention describes a method for transforming a high  
CC temperature-resistant coryneform microbe in which the high temperature-  
CC resistant coryneform microbe is treated with a chemical changing the  
CC structure of cell wall of the microbe and an electric pulse is applied  
CC to a solution containing the microbe having a changed structure of cell  
CC wall and a DNA. Also described are: (1) transforming a high temperature-  
CC resistant coryneform microbe in which the cell wall of a high  
CC temperature-resistant coryneform microbe is weakened and electric pulse  
CC is applied to a solution containing the microbe having weakened cell  
CC wall and a DNA; (2) a transformant of a high temperature-resistant  
CC coryneform microbe in which a recombinant DNA is introduced; and  
CC (3) a transformant of a high temperature-resistant coryneform microbe  
CC carrying both a plasmid vector derived from the high temperature-  
CC resistant coryneform microbe and a plasmid vector derived from  
CC Corynebacterium glutamicum or Brevibacterium lactofermentum. The method  
CC is used for transforming a high temperature-resistant coryneform microbe.  
CC The present sequence encodes a Brevibacterium lactofermentum pSFK6  
CC protein sequence, which is used in an example from the present  
CC invention.

SQ Sequence 4447 BP; 824 A; 997 C; 1378 G; 1248 T; 0 other;

Query Match 3.3%; Score 37.4; DB 24; Length 4447;

Best Local Similarity 53.0%; Pred. No. 0.87; Indels 0; Gaps 0;

Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 493 AACGGCTGCGACGTCACGCTTCACTGATGTTGAGTGACGAGGAGAAC 552  
DB 1795 AACGGTTGCACTTGACGACGCAACATGCTGTGTTGATGATGTCATCTGACGAT 1054

QY 553 ATCCGGAATCCTTCTCGAGATGCGATTCGATGAGATTCGAACTGATCTCG 612  
DB 1855 GAACTCAGGCGGCTTGAGGATTCAGTTTTCCCGCTGCTGCTGTGTTAAGCC 1914

QY 613 GGATTTCGCCCACTAGTATTCGGG 643

DB 1915 GGATGACGCGCCACTGCGTGACGCGG 1945

## RESULT 9

ABL49734

ID ABL49734 standard; DNA; 4447 BP.

AC ABL49734;

DT 29-MAY-2002 (first entry)

DE Brevibacterium lactofermentum p48K nucleotide sequence SEQ ID NO:7.

KM Kanamycin resistant gene; mutation; high temperature resistance;

KM coryneform microbe; transformation; gene; ds.

OS Brevibacterium lactofermentum.

FH Key Location/Qualifiers

FT CDS 1318..2601

FT /\*tag= a

FT /product= "p48K"

PN JP2002017362-A.

PD 22-JAN-2002.

PF 05-JUL-2000; 2000JP-0204236.

PR 05-JUL-2000; 2000JP-0204236.

PA (AJIN ) AJINOMOTO KK.

DR MPI; 2002-263243/31.

DR P-PSDB; ABB06342.

PT Transformation of high temperature-resistant coryneform microbe and

PT transformant -

PS Example; Page 20-22; 28pp; Japanese.

CC The present invention describes a method for transforming a high  
CC temperature-resistant coryneform microbe in which the high temperature-  
CC resistant coryneform microbe is treated with a chemical changing the  
CC structure of cell wall of the microbe and an electric pulse is applied  
CC to a solution containing the microbe having a changed structure of cell  
CC wall and a DNA. Also described are: (1) transforming a high temperature-  
CC resistant coryneform microbe in which the cell wall of a high  
CC temperature-resistant coryneform microbe is weakened and electric pulse  
CC is applied to a solution containing the microbe having weakened cell  
CC wall and a DNA; (2) a transformant of a high temperature-resistant  
CC coryneform microbe in which a recombinant DNA is introduced; and  
CC (3) a transformant of a high temperature-resistant coryneform microbe  
CC carrying both a plasmid vector derived from the high temperature-  
CC resistant coryneform microbe and a plasmid vector derived from  
CC Corynebacterium glutamicum or Brevibacterium lactofermentum. The method  
CC is used for transforming a high temperature-resistant coryneform microbe.  
CC The present sequence encodes a Brevibacterium lactofermentum p48K  
CC protein sequence, which is used in an example from the present  
CC invention.

SQ Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;

Query Match 3.3%; Score 37.4; DB 24; Length 4447;

Best Local Similarity 53.0%; Pred. No. 0.87; Indels 0; Gaps 0;

Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;







CC compared with 20, mg/dl. The present sequence is  
CC Brevibacterium lactofermentum p48K replication control region DNA.  
CC This DNA is used in the construction of shuttle vector for  
CC Escherichia coli and Corynebacterium bacteria and temperature sensitive  
CC vector.  
SQ Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;  
Query Match 3.3%; Score 37.4; DB 24; Length 4447;  
Best Local Similarity 53.0%; Pred. No. 0.87;  
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 493 AACGCTGGCAGCCGCTTACGCGCTACTCATGTTCAAGTGGTGAACGTGAAGAAC 552  
DB 1795 AACGTTGGACATTGCAACCGCAACATGCTGTTCTTGATCGTCCACTGTGACGAT 1854  
QY 553 ATCTCGAATCTCTCTCGATGCGATGTCGATCGTGAAGCTTCAAACTCGATCTCTG 612  
DB 1855 GAACTCAAGGATTTGAGGANTGCATGTTTCCGCGTGTCTGTGTGTTAAAGCC 1914  
QY 613 GGATTGCTGGCGCCACTACGTAATTCGAGTG 643  
DB 1915 GGTATGACGCGCCACTGCGTGAACGCGGG 1945  
RESULT 12  
AAH24065/c  
ID AAH24065 standard; DNA; 1631 BP.  
AC AAH24065;  
XX  
DT 05-OCT-2000 (first entry)  
XX  
DE E. coli proliferation associated coding sequence SEQ ID NO:208.  
XX  
KM Escherichia coli; E. coli; proliferation; inhibition; screening;  
KM antimicrobial; bacterial growth; antisense therapy; antibacterial; ds.  
XX  
OS Escherichia coli.  
XX  
PN WO200044906-A2.  
XX  
PD 03-AUG-2000.  
XX  
PF 27-JAN-2000; 2000WO-US02200.  
XX  
PR 27-JAN-1999; 99US-0117405.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GT;  
PI Yamamoto RT, Xu HH;  
XX  
XX WPI: 2000-514822/46.  
DR P-PSDB; AAB16010.  
XX  
XX Novel polynucleotides and polypeptides associated with microorganism  
PT proliferation, used to identify inhibitors of bacterial growth and  
PT proliferation, for use in antisense therapy -  
XX  
PS Claim 8; Page 162-163; 316pp; English.  
XX  
XX AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide  
CC sequences derived from Escherichia coli which inhibit E. coli  
CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent  
CC nucleotide and protein sequences associated with E. coli proliferation.  
CC AAA66056 and AAA66057 represent primers used for sequencing E. coli  
CC proliferation inhibiting nucleotide inserts in an example from the  
CC present invention. Methods from the present invention can be used to  
CC identify a proliferation-regulated gene in a microorganism, by contacting  
CC a microorganism with a proliferation-regulated gene activity inhibitory  
CC nucleic acid identified in another organism, and determining if  
CC inhibition occurs in the second microorganism. The nucleic acid sequences

CC identified as being required for bacterial growth and proliferation, can  
CC be used for antisense therapy for killing bacteria.  
XX  
SQ Sequence 1631 BP; 463 A; 340 C; 429 G; 399 T; 0 other;  
Query Match 3.1%; Score 34.8; DB 21; Length 1631;  
Best Local Similarity 50.0%; Pred. No. 3.5;  
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
QY 219 CGAAGAGGCTGATGTCGCCCTGCTGTGCGGAAAGTCGTGCACATGTCAGACGA 278  
DB 1184 CGAAGATGCTGAGAGGACCGAGATGGCAAAAAAATGCGCGAAGCTATGCTGCC 1243  
QY 279 AATTCTCAAGTTGTCCTCATCACTCGGACTGATCTGTTCCATGTCAGATGAC 338  
DB 1244 ATTATCAAGAAGATGAATGCTCAATTGCTGCTCGAATTAATGAACAGGGCGAATAT 1303  
QY 339 CATGGCCATACAGCTGGTGTAGAGGGCTCCAGACCTATGATGATGATTTCCGC 392  
DB 1304 AGCCCGTATGAGCTGTTAAGATGCGCAAGAAATCTGTGAATTTCTGC 1357  
RESULT 13  
AAH24065/c  
ID AAH24065 standard; DNA; 4590 BP.  
AC AAH24065;  
XX  
DT 29-AUG-2001 (first entry)  
XX  
DE Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.  
XX  
XX Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;  
KM modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;  
KM lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;  
KM functional food; transgenic yeast; fat/lean ratio; food use; ds.  
XX  
OS Saccharomyces cerevisiae.  
XX  
FH Location/Qualifiers  
FH 10  
FT misc\_feature  
FT /tag= a  
FT /note= "Represented as \* in the specification"  
FT 3617  
FT /tag= b  
FT /note= "Represented as \* in the specification"  
FT 3649  
FT misc\_feature  
FT /tag= c  
FT /note= "Represented as \* in the specification"  
FT 3679  
FT /tag= d  
FT /note= "Represented as \* in the specification"  
FT 3819  
FT misc\_feature  
FT /tag= e  
FT /note= "Represented as \* in the specification"  
FT 3862  
FT /tag= f  
FT /note= "Represented as \* in the specification"  
FT 3864  
FT misc\_feature  
FT /tag= g  
FT /note= "Represented as \* in the specification"  
FT 3888  
FT /tag= h  
FT /note= "Represented as \* in the specification"  
FT 3890  
FT misc\_feature  
FT /tag= i  
FT /note= "Represented as \* in the specification"  
FT 3912  
FT /tag= j  
FT /note= "Represented as \* in the specification"  
FT 3914  
FT misc\_feature  
FT /tag= k  
FT /note= "Represented as \* in the specification"

```
FT misc_feature 3938
FT /*tag= 1
FT /note= "Represented as * in the specification"
FT misc_feature 3939
FT /*tag= m
FT /note= "Represented as * in the specification"
FT misc_feature 3941
FT /*tag= o
FT /note= "Represented as * in the specification"
FT misc_feature 3943
FT /*tag= p
FT /note= "Represented as * in the specification"
FT misc_feature 4361
FT /*tag= q
FT /note= "Represented as * in the specification"
XX
```

```
PN W0200133977-A1.
XX
XX 17-MAY-2001.
XX
XX 06-NOV-2000; 2000WO-AU01362.
XX
XX 05-NOV-1999; 99AU-0003875.
XX
XX (META-) METABOLIC PHARM LTD.
XX
XX Belyea CI, Ng FM, Vaughan P;
XX
XX WPI; 2001-328876/34.
XX
XX New organisms containing nucleic acid encoding a growth hormone
XX fragment which modulates lipid metabolism are useful to produce dietary
XX aids for obesity and in the meat production industry
XX
XX Disclosure; Page 48-50; 54pp; English.
```

```
XX
XX The invention relates to novel transgenic organisms useful in the
XX production of functional food and drink products for the treatment
XX or prevention of obesity via the regulation of lipid metabolism. The
XX organisms comprise a polynucleotide encoding a growth hormone fragment
XX capable of stimulating the activity of hormone-sensitive lipase (the key
XX enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
XX enzyme in lipogenesis). The growth hormone fragment preferably contains
XX at least the disulphide-bonded loop of a mammalian growth hormone (but is
XX not the full-length growth hormone) and is optionally linked to an
XX epitope tag or heterologous fusion protein partner. The transgenic
XX organism may be a microorganism used to produce a fermented product
XX (e.g., yeast), or an edible plant or animal or cell thereof. Food or
XX drink made using methods of the invention are used to modify fat/lean
XX ratio, lipid metabolism or food use in a mammal. In particular, the food
XX or drink products may be used to treat or prevent obesity, particularly
XX in humans, and may also be used to improve the fat/lean ration of
XX livestock raised for meat production. In the exemplification of the
XX invention, the human growth hormone (hGH) fragment analogue A0D9604 was
XX expressed in yeast, optionally fused to the FLAG epitope (AAB73625).
XX The present sequence is described as a DNA sequence from yeast in
XX the sequence listing, but is not further referred to in the
XX specification.
```

```
XX SQ Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;
```

```
Query Match 3.1%; Score 34.8; DB 22; Length 4590;
Best local similarity 13.2%; Pred. No. 5.9;
Matches 75; Conservative 211; Mismatches 201; Indels 2; Gaps 1;
```

```
QY 455 ACGGATACGTGCGCGCTGTTGAATCACTACGAGAAAACGCTGGACGCTCCAGCTTC 514
Db 765 AMGNMTDAMGKSHKSTGSKASRSBYTVDYHNTTVAAGDASSTSHATVNCCHYNXYTA 706
QY 515 ACGGCTACTCATTTGAGTGTGACGTGAGTGAAGCAATCCTCGAATCCTTCGGATG 574
Db 705 HYDYSKNSNAGNANCGHNTYAKYRAVYTTDGDWSATCNGNCRMGTAHSHSTGCTYKHK 646
```

```
QY 575 CGATGTTGATCGGTGAGCTTCCAACTGTAATCTGGAATTTGCTGGCCACTACGTA 634
Db 645 YSNYNAGAAVYTTDDMDMATCGMGYASSHKYNHTTKDQSHKTKNTNNSNAGNBANCGH 586
QY 635 ATTCCGGTGTCTCCGATGACGAAAGATGGCGGCTGAACCTGATCAAGTTCTGCTGCT 694
Db 585 NTYAKNGHATYKTY--HBTAVYTTDDWSATCNGKTNHYGSHSTKYANTCNSTGTAN 528
QY 695 ATCTGACGAAAATTGATGCTGGCTGTGATGAGAGTTGATGGCGGAGGAAAAGTG 754
Db 527 KHTDAMHATYASKNTNMSTAHGTRSBTNNAGTYKAVYTTDDWSATCNGSTWSTGT 468
QY 755 GTCCACATGCAACCGGTGACCCCTGGAAATGCTGTGATGACGATGCGCGGATCCAC 814
Db 467 SSTNAGTNTSYVASRSRTBTYNTSYRASRSRTTBYBASRSRTNSTYRASRSYRAS 408
QY 815 AAGCGTTGAACTGTGCGCGGAAATTTGATGTTGCTTCATGGAGCTGGGCAATGCGT 874
Db 407 RSNSTYRASRYRASNSNSTSRNSSTYRAVRAVNSTSTHVNANTBGTNSYRAVNSTY 348
QY 875 GGTCCCGTGAATGCGTGGCCCGAGCTGCTTGGGGCAGAACTAACGATGCTCAGATCG 934
Db 347 NSTYNSYTRAYNSYNSYNSYTRAYNSYNSYNSYNSYNSYNSYNSYNSYNSYNSY 288
QY 935 TTGAGCAGAAAGAAATCTGCCCGGTGATGTCGATTCATTCGCGCGCATCGGATGA 994
Db 287 TDTGVVDYHRTDHTSYTRAYNSNTTRNDKTHSYDAYVDYRAHYRASNYSHSRSNY 228
QY 995 TGATTCGACTTGTGGCGCTTACGCTTC 1023
Db 227 SRTTBSRTBTTSRTTSRTTSRTTSRTTB 199
```

## RESULT 14

```
AAQ55750
ID AAQ55750 standard; DNA; 1505 BP.
```

```
AC AAQ55750;
XX
DT 25-MAR-2003 (updated)
DT 23-JUL-1994 (first entry)
XX
DE Genomic clone g11F includes a Brassica root specific promoter.
XX
KM Plants; transformation; food crops; toxins; insects; fungal;
XX infection; ss.
XX
OS Brassica napus.
XX
XX
XX Key Location/Qualifiers
XX FT misc_feature 552..554
XX /*tag= a
XX FT /note= "translation start site"
XX FT CAAAT_signal 438..440
XX /*tag= b
XX FT TATA_signal 485..488
XX /*tag= c
XX
```

```
PN W09402619-A1.
```

```
XX 03-FEB-1994.
```

```
PF 12-JUL-1993; 93WO-US06541.
```

```
XX 16-JUL-1992; 92US-0915246.
```

```
PA (PION-) PIONEER HI-BRED INT INC.
```

```
PI Baszczynski CJ, Bellemare G, Boivin R, Fallis L;
```

```
XX WPI; 1994-048877/06.
```

```
XX Brassica root-specific promoter sequence and vectors containing
```

PT it - for expression of e.g. toxins in the roots of crops  
PT susceptible to root diseases  
XX Disclosure; Fig 4; 54bp; English.  
XX  
CC The sequence shows the G1F genomic clone isolated from DNA from  
CC the roots of Brassica napus and containing a promoter sequence  
CC upstream from the translation start site. The promoter enhances  
CC gene expression in roots and can be used to confer disease  
CC resistance or immunity upon a susceptible plant root by transforming  
CC the plant with a composite gene capable of abundant expression in the  
CC gene, comprising the promoter and the gene of interest. This method  
CC can be used to transform plants, partic food crops that have roots  
CC susceptible to fungal or insect diseases. The roots of the transformed  
CC plants express a toxin or substance that provides immunity or resistance  
CC to the fungal or insect diseases.  
CC See also AA055749 and AA056913-5.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 1505 BP; 395 A; 201 C; 528 G; 381 T; 0 other;  
Query Match 3.0%; Score 34.6; DB 15; Length 1505;  
Best Local Similarity 58.1%; Pred. No. 3.9; Indels 0; Gaps 0;  
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
QY 706 ATTGCATCTGCGCTGTGATGAGGTGAGTGTGCGGACGAAAAGTGTGACATGCG 765  
DB 699 ATTGAGTTGATGCTGTGTGTGATGAGGTGCGGACGAAAAGTGTGCTGTGT 758  
QY 766 AACCGTCACCCCTGGGAAATCGTGTGATGACATGCGGCGGAT 810  
DB 759 TACGGTGAAGCTGAAGACATGTGTGAGAGAGAGCGGCTGACAT 803  
RESULT 15  
ABI99802/c  
ID ABI99802 standard; cDNA; 1954 BP.  
XX  
AC ABI99802;  
XX  
DT 07-MAR-2002 (first entry)  
XX  
DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:908.  
XX  
KM Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
OS Mus musculus.  
PN WO200188188-A2.  
XX  
PD 22-NOV-2001.  
XX  
PF 18-MAY-2001; 2001WO-0P04192.  
XX  
PR 18-MAY-2000; 2000JP-0145977.  
XX  
PA (UNINT-) UNIV NITHON SCHOOL JURIDICAL PERSON.  
XX  
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
XX  
DR WPT; 2002-034733/04.  
XX P-PSDB; ABB57327.  
XX  
PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
PT expression levels of particular genes defined in the specification or  
PT by determining the expression profile of a gene group comprising these  
PT genes -  
XX  
PS Claim 2; Page 2294-2297; 2690pp; English.  
XX  
CC The present invention describes a method for examining ischaemic  
CC conditions, comprising measuring the expression levels of particular

CC genes (I) in a test sample or determining the expression profile of a  
CC gene group in the sample comprising genes selected from (I). The method  
CC is useful for examining the ischaemic condition (e.g. compressive  
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
CC expression levels of particular genes (ABI99202 to ABI99912, encoding  
CC the protein sequences in ABB57020 to ABB57374) or by determining the  
CC expression profile of a gene group comprising these genes. The  
CC expression levels or expression profiles produced by these genes are  
CC used as an indicator when screening for ischaemic condition-improving  
CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914  
CC represent PCR primers for a mouse ischaemic condition related sequence,  
CC which are used in the exemplification of the present invention.  
XX  
SQ Sequence 1954 BP; 451 A; 599 C; 550 G; 353 T; 1 other;  
Query Match 3.0%; Score 34.6; DB 24; Length 1954;  
Best Local Similarity 47.8%; Pred. No. 4.5; Indels 0; Gaps 0;  
Matches 100; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
QY 609 TCTGGGATTGCTGCGCCACTACGTAATTCGGGTGCTCGATGTAAGAAAGTCGGCG 668  
DB 1213 TGTGGGTGCTGCTGCGCGGAGAGATGAGATGCGGATGCGGATGAGAGTGAAC 1154  
QY 669 TGAAGCTATCAAGTTCCTGCGCGTATCTGACAAAATTGATCTGGGGTTGATGGA 728  
DB 1153 TGTTCGGGTTGGGGCTGAGAGTTGGGGCTGAGGTTGGGGCTGGGGTTGA 1094  
QY 729 GGTGTGATGTCGACGAGAAAAGTGTGACATGCGCAACCTGCAACCTGGGAAATCGC 788  
DB 1093 GGTGTGATGTCGACGAGGCTGGGGCTGGGGTTGATCTCGGCTGGGGCTGGGATCTCC 1034  
QY 789 TGTTCATGTCAGTGGGCGGGGATCCAG 817  
DB 1033 TCGCTCTGCGGAGGCTGCTGTGACCAAG 1005

Search completed: November 8, 2003, 20:05:48  
Job time : 284 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 19:20:06 ; Search time 69 Seconds

(without alignments)  
7292.417 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1140  
Sequence: 1 atgaccagcgaagtcgctga.....taatcgcgtcgcaagtcga 1140

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/ECTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37.4	3.3	4447	4 US-09-521-668B-17	Sequence 17, Appl
2	37.4	3.3	4447	4 US-09-521-668B-19	Sequence 19, Appl
3	34.6	3.0	1505	1 US-07-915-246-1	Sequence 1, Appl
4	34.4	3.0	7218	1 US-08-232-463-14	Sequence 14, Appl
5	33.6	2.9	1134	4 US-09-482-180A-3	Sequence 3, Appl
6	33.2	2.9	1780	3 US-08-813-150-3	Sequence 3, Appl
7	33.2	2.9	1780	4 US-09-546-553-3	Sequence 3, Appl
8	33.2	2.9	1867	4 US-09-482-273-81	Sequence 81, Appl
9	32.8	2.9	1987	1 US-08-722-001-26	Sequence 26, Appl
10	32.8	2.9	2793	1 US-08-209-747-1	Sequence 1, Appl
11	32.8	2.9	2793	1 US-08-458-298-1	Sequence 1, Appl
12	31.6	2.8	1356	4 US-09-252-991A-14918	Sequence 14918, A
13	31.4	2.8	4403765	3 US-09-103-840A-2	Sequence 2, Appl
14	31.4	2.8	4411529	3 US-09-103-840A-1	Sequence 1, Appl
15	31.2	2.7	1599	4 US-09-252-991A-13921	Sequence 13921, A
16	31.2	2.7	2280	4 US-09-252-991A-13535	Sequence 13535, A
17	31.2	2.7	5741	1 US-07-706-699-4	Sequence 4, Appl
18	31.2	2.7	5741	1 US-07-998-931-4	Sequence 4, Appl
19	30.8	2.7	5820	3 US-09-029-213B-7	Sequence 7, Appl
20	30.6	2.7	1708	5 PCT-US95-02481-3	Sequence 3, Appl
21	30.6	2.7	4403765	3 US-09-103-840A-2	Sequence 2, Appl
22	30.6	2.7	4411529	3 US-09-103-840A-1	Sequence 1, Appl
23	30.4	2.7	1001	4 US-09-641-638-315	Sequence 315, App
24	30.4	2.7	2128	4 US-09-280-116-11	Sequence 11, Appl
25	30.4	2.7	2459	4 US-09-443-795-2	Sequence 2, Appl
26	30.4	2.7	3353	3 US-09-532-803-10	Sequence 10, Appl
27	30.4	2.7	3353	4 US-09-653-403-1	Sequence 1, Appl

## ALIGNMENTS

28	30.4	2.7	3333	4 US-10-013-784-1	Sequence 1, Appl
29	30.2	2.6	11707	3 US-09-136-574A-1	Sequence 1, Appl
30	30.2	2.6	18596	3 US-09-318-448-11	Sequence 11, Appl
31	30.2	2.6	18597	4 US-09-962-665-8	Sequence 8, Appl
32	30	2.6	289	3 US-09-007-005-17	Sequence 17, Appl
33	30	2.6	289	3 US-09-244-796-17	Sequence 17, Appl
34	30	2.6	504	4 US-09-252-991A-12357	Sequence 12357, A
35	30	2.6	1200	4 US-09-252-991A-12403	Sequence 12403, A
36	30	2.6	1232	4 US-09-462-645C-3	Sequence 3, Appl
37	30	2.6	1233	4 US-09-252-991A-12188	Sequence 12188, A
38	30	2.6	2949	4 US-09-023-905A-8	Sequence 8, Appl
39	30	2.6	3481	4 US-09-462-645C-11	Sequence 11, Appl
40	30	2.6	3810	4 US-09-462-645C-7	Sequence 7, Appl
41	30	2.6	5954	4 US-09-023-905A-6	Sequence 6, Appl
42	30	2.6	36470	4 US-08-311-731A-123	Sequence 123, App
43	29.8	2.6	1353	4 US-09-252-991A-9189	Sequence 9189, Ap
44	29.8	2.6	1365	4 US-09-252-991A-13442	Sequence 13442, A
45	29.8	2.6	1464	4 US-09-252-991A-8876	Sequence 8876, Ap

## RESULT 1

US-09-521-668B-17  
; Sequence 17, Application US/09521668B  
; Patent No. 6303383  
; GENERAL INFORMATION:  
; APPLICANT: NAKAMURA, JUN  
; APPLICANT: KANNO, SOHEI  
; APPLICANT: KIMURA, KATSUMI  
; APPLICANT: MATSUI, KAZUHIRO  
; APPLICANT: NAKAMATSU, TOSYOSHI  
; TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA  
; FILE REFERENCE: 0010-1093-0  
; CURRENT APPLICATION NUMBER: US/09/521,668B  
; CURRENT FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: JP 11-69896  
; PRIOR FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 17  
; LENGTH: 4447  
; TYPE: DNA  
; ORGANISM: Brevibacterium lactofermentum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1316)..(2598)  
US-09-521-668B-17

## Query Match

Best Local Similarity 53.0%; Pred. No. 0.18;  
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 493 AACGGTCGACGTCACGCTTACGAGCGGCTACGACGTCGAGTCGAGTCGAGC 552  
DB 1795 AACGGTCGACGTCGACGCTTACGAGCGGCTACGAGTCGAGTCGAGTCGAGC 1854  
QY 553 AACGTCGATTCCTTCGATGCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 612  
DB 1855 GAACCTAAGCGCTTGAAGATTCATGTTCCCGCTGCTGCTGCTGCTGCTGAAGCC 1914  
QY 613 GGATTCGTCGCGCCTACGTCGATTCGAGGTCG 643  
DB 1915 GGTATGAGCGCGCCTACGTCGATTCGAGGTCG 1945

## RESULT 2

US-09-521-668B-19  
; Sequence 19, Application US/09521668B  
; Patent No. 6303383  
; GENERAL INFORMATION:  
; APPLICANT: NAKAMURA, JUN

APPLICANT: KANNO, SOHEI  
APPLICANT: KIMURA, EIICHIRO  
APPLICANT: MATSUI, KAUIKO  
APPLICANT: NAKAMATSU, TSUYOSHI  
TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA  
FILE REFERENCE: 0010-1093-0  
CURRENT APPLICATION NUMBER: US/09/521,668B  
CURRENT FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: JP 11-69896  
PRIOR FILING DATE: 1999-03-16  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 19  
LENGTH: 4447  
TYPE: DNA  
ORGANISM: Brevibacterium lactofermentum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1318)..(2598)  
US-09-521-668B-19

Query Match 3.3%; Score 37.4; DB 4; Length 4447;  
Best Local Similarity 53.0%; Pred. No. 0.18; Indels 0; Gaps 0;  
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 493 AACGCTGACCTCCACGTTCCAGCGCTACTCATGTTCACTGAGTGAAGAAC 552  
DB 1795 AACGTTGGACCTTGACCGCCACATGCTGTTCTTGATGTCACGTCGTGAGAT 1854  
QY 553 ATCCTGATCTCTTCGGATGAGTGTGATGCGTGGTGAACCTGATCTCTG 612  
DB 1855 GAACCTAGGACATTTGAGATTCATGTTCCCGCTGCTGCTGTTAAAGGCC 1914  
QY 613 GGATTTGCTGGCGCCACAGCTAATTCGGGTG 643  
DB 1915 GGATGAGCGCGCCACTGCGTGAAGCAGCGG 1945

## RESULT 3

US-07-915-246-1  
Sequence 1, Application US/07915246  
Patent No. 5401836

GENERAL INFORMATION:  
APPLICANT: Basczynski, Chris L.  
APPLICANT: Fallis, Lynne  
APPLICANT: Bellmare, Guy  
APPLICANT: Boivin, Rodolphe  
TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAndrews, Held, and Malloy  
STREET: 500 W. Madison St. 34th Floor  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60661  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/915,246  
FILING DATE: 19920716  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pochopien, Donald J.  
REGISTRATION NUMBER: 32,167  
REFERENCE/DOCKET NUMBER: 91 P 1125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312 707-8889  
TELEFAX: 312 707-9155

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1505 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLSCULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Brassica napus  
STRAIN: Westar  
DEVELOPMENTAL STAGE: Somatic  
TISSUE TYPE: Root  
US-07-915-246-1

Query Match 3.0%; Score 34.6; DB 1; Length 1505;  
Best Local Similarity 58.1%; Pred. No. 0.8;  
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 706 ATTGATCTGCGCTTGATGAGGTTGTGAGCGAAGAAAAAGTGTGACATGCG 765  
DB 699 ATTGAGTTGATGCTGCTGTTGTTGTTGAGTGTGCGGAGAGAGTGTGCTGCT 758  
QY 766 AACGTCACCTCGGAAATCGCTGTTGATGCACTGGCGCGGAT 810  
DB 759 TACGGTGAGCTGAAGACATGTTGAGAGAGAGAGCGCGTGACAT 803

## RESULT 4

US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367

GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHREIFLINGER, F.  
TITLE OF INVENTION: RECOMBINANT FOMLPD VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single



Db 1076 CCGCCACATGCAGAGAGCGCTGCGAGGCTGCACTACGTGCCCGCCACCAAGGTGTTTC 1133

## RESULT 7

US-09-546-553-3

Sequence 3, Application US/09546553

Patent No. 6518405

GENERAL INFORMATION:

APPLICANT: Mueller, Christopher

APPLICANT: Lebecque, Serge J.E.

APPLICANT: Liu, Yong-Jun

APPLICANT: Dowling, Lynette M.

APPLICANT: Huffine, Constance M.

APPLICANT: Gorman, Daniel M.

TITLE OF INVENTION: MAMMALIAN PROTEINASES; OXIDOREDUCTASES;

TITLE OF INVENTION: RELATED REAGENTS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: DNAX Research Institute

STREET: 901 California Avenue

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/546,553

FILING DATE: 10-APR-2000

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/813,150

FILING DATE: 07-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Ching, Edwin P.

REGISTRATION NUMBER: 34,090

REFERENCE/DOCKET NUMBER: SF0693

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-852-9196

TELEFAX: 650-496-1200

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1780 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 66..1766

US-09-546-553-3

Query Match 2.9%; Score 33.2; DB 4; Length 1780;

Best Local Similarity 55.1%; Pred. No. 2.4;

Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Db 640 GGTGCTCGATGTACGAAAGATCGCGTGAAGCTGATCAAGTTCTGCGTATCTG 699  
1016 GGTGCTCGATGTACGCGGACCGGACCGGCGGTGAAGCGCATCACTTCTGCGCGCTGCC 1075  
QY 700 ACGAAATTCATCTGCGGTGTGTATGAGGTTGTAGTGGCGACGAAAGTGTGTC 757  
Db 1076 CCGCCACATGCAGAGAGCGCTGCGAGGCTGCACTACGTGCCCGCCACCAAGGTGTTTC 1133

## RESULT 8

US-09-482-273-81

Sequence 81, Application US/09482273

Patent No. 6534631

GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 71 Human Secreted Proteins  
FILE REFERENCE: P2030P1

CURRENT APPLICATION NUMBER: US/09/482,273

CURRENT FILING DATE: 2000-01-13

EARLIER APPLICATION NUMBER: PCT/US99/15849

EARLIER FILING DATE: 1999-07-14

EARLIER APPLICATION NUMBER: 60/092,921

EARLIER FILING DATE: 1998-07-15

EARLIER APPLICATION NUMBER: 60/092,922

EARLIER FILING DATE: 1998-07-15

EARLIER APPLICATION NUMBER: 60/092,956

EARLIER FILING DATE: 1998-07-15

NUMBER OF SEQ ID NOS: 267

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 81

LENGTH: 1867

TYPE: DNA

ORGANISM: Homo sapiens

US-09-482-273-81

Query Match 2.9%; Score 33.2; DB 4; Length 1867;

Best Local Similarity 55.1%; Pred. No. 2.5;

Matches 65; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 640 GGTGCTCGATGTACGAAAGATCGCGTGAAGCTGATCAAGTTCTGCGTATCTG 699  
Db 1053 GGTGCTCGATGTACGCGGACCGGACCGGCGGTGAAGCGCATCACTTCTGCGCGCTGCC 1112  
QY 700 ACGAAATTCATCTGCGGTGTGTATGAGGTTGTAGTGGCGACGAAAGTGTGTC 757  
Db 1113 CCGCCACATGCAGAGAGCGCTGCGAGGCTGCACTACGTGCCCGCCACCAAGGTGTTTC 1170

## RESULT 9

US-08-722-001-26

Sequence 26, Application US/08722001

Patent No. 5760054

GENERAL INFORMATION:

APPLICANT: Thompson, Wayne J.

APPLICANT: Huff, Joel R.

APPLICANT: Nerenderg, Jennie B.

APPLICANT: Lee, Hee-Yoon

TITLE OF INVENTION: ALPHALIC ADRENERGIC RECEPTOR ANTAGONISTS

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.

STREET: 126 Lincoln Avenue

CITY: Rahway

STATE: New Jersey

COUNTRY: United States of America

ZIP: 07065

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/722,001

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/229,276

FILING DATE: 14-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Appollina, Mary A.

REGISTRATION NUMBER: 34,087

REFERENCE/DOCKET NUMBER: 19169Y

TELECOMMUNICATION INFORMATION:

TELEPHONE: (908) 594-3462

TELEFAX: (908) 594-4720



TELEX: 138825  
INFORMATION FOR SEQ ID NO: 26;  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1987 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-722-001-26

Query Match 2.9%; Score 32.8; DB 1; Length 1987;  
Best Local Similarity 56.5%; Pred. No. 3.4;  
Matches 61; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 105 ACAATCACCACGCTCAAGAAACATTAAAGCCTGTGGCCGCCGATTTGCGCTGAACG 164  
DB 778 ACCGACTACTACATCGTCGACACCGCTAGTGGCGTGGCCGACCTCCGCTCACTCCACG 837  
QY 165 TGTGACCATGTCAAGGCTCCGAAAGGTTTGGATTGCGAGGCTTCG 212  
DB 838 TGTGCTCCCTCTCCGCCATCTTCAGAGTCTTAGGCTACTGGGCTTCG 885

RESULT 10  
US-08-209-747-1  
Sequence 1, Application US/08209747  
Patent No. 5733771  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider  
TITLE OF INVENTION: Silk Proteins  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Birch, Stewart, Kolaesch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-3487  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/209,747  
FILING DATE: 14-MAR-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-104P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Nephila clavipes  
TISSUE TYPE: minor ampullate gland  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 183..2675  
OTHER INFORMATION: /product= "N. clavipes minor

OTHER INFORMATION: ampullate silk protein  
US-08-209-747-1

Query Match 2.9%; Score 32.8; DB 1; Length 2793;  
Best Local Similarity 39.8%; Pred. No. 4.1;  
Matches 82; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 708 TGCATGTGCGCTTGTGATAGAGTTGTGTGCGGACGAAAAAGTGTGACATGCGAA 767  
DB 899 TGAAGCTGAGCGCTGACGAGTGCAGAGACAGAGCTGAACTACGCTGTCAAGATA 958  
QY 768 CCGTCAACCTGGGAAATGCTGTGTGATGAGGCGGAGATCCACAAGCTTGAAC 827  
DB 959 CGGTGCCGAGACGAGAGCTGTGTCTCTGCTGACGCTGANNNNNNNNNNNNNNNN 1018  
QY 828 GTGGCAGAAATTGACTTTGTTCATGAGACGTCGCGCAATCGCTGTCCCGTGAAT 887  
DB 1019 NNN 1078  
QY 888 GCGTGCCGAGCTGTCTTTGGGCGAG 913  
DB 1079 TGTGTCGAGAGCAGAGCTGTGTGCGG 1104

RESULT 11  
US-08-458-298-1  
Sequence 1, Application US/08458298  
Patent No. 5756677  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider  
TITLE OF INVENTION: Silk Proteins  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Birch, Stewart, Kolaesch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-3487  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,298  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/209,747  
FILING DATE: 14-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-104P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Nephila clavipes  
TISSUE TYPE: minor ampullate gland  
FEATURE:  
NAME/KEY: CDS



QY	411	AAAGGTCGTGGTGGGTACGAAACGTGAATGACGGCTGGCAAGATACGTGGCGC	470
Db	3360971	CCACGATGATCCGTGACTGTGCTTGCACTGGTGGCCACCAACGCCAAGCACTCGA	3361030
QY	471	TGTTGAATCACTCACGGAAGAAAACGGCTGGACAGTCACGTTTACAGCGCTACTCATGTT	530
Db	3361031	TGTTAAACCGCGCCGGAGAACAGCGCCGACCCCGGCAGACAGCGCGGCTTGTCTT	3361090
QY	531	C 531	
Db	3361091	C 3361091	

## RESULT 15

```

US-09-252-991A-13921
: Sequence 13921, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,768
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 13921
: LENGTH: 1599
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13921

```

	Query Match	Similarity	2.7%	Score 31.2	DB 4	Length 1599
	Best Local	Similarity	57.0%	Pred. No. 9.7		
	Matches	Conservative	0	Mismatches 43	Indels 0	Gaps 0
QY	471	TGTTGAATCACTACCGGAAAAACGGCTGGCAGTCCAGCTACCGGCTACATCATT	530			
Db	1287	TGTTGACTTCGGCTTCGATTACGCTCGGAGGACCCACAGCTCATGGCGGCTCAAGAG	1346			
QY	531	CAGTGTGACGCTGAGTGAAGACATCCTGAAATCCTCTTCG	570			
Db	1347	CGAAAGTCTTTTCAACAGAACTTCATGATATTTCTTCGG	1386			

Search completed: November 8, 2003, 21:29:58  
Job time : 85 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 20:01:41 / Search time 304 Seconds  
(without alignments)

11966.092 Million cell updates/sec

Title: US-10-007-527A-1

Sequence: 1 atgaccagcgtaagtcgta.....taatacgttcgtaagtcga 1140

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2141354 seqs, 1595478879 residues

Total number of hits satisfying chosen parameters: 4282708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1140	100.0	1140	US-10-007-527A-1	Sequence 1, Appli
2	1140	100.0	1140	US-10-007-452-1	Sequence 1, Appli
3	1140	100.0	6334	US-10-007-527A-5	Sequence 5, Appli
4	1140	100.0	6334	US-10-007-452-5	Sequence 5, Appli
5	1140	100.0	9652	US-10-007-527A-7	Sequence 7, Appli
6	1140	100.0	9652	US-10-007-452-7	Sequence 7, Appli
7	1140	100.0	11241	US-10-007-527A-6	Sequence 6, Appli
8	1140	100.0	11241	US-10-007-452-6	Sequence 6, Appli
9	40.8	3.6	520	US-10-184-644-332	Sequence 332, App
10	40.8	3.6	520	US-10-184-634-332	Sequence 332, App
11	37.6	3.3	594	US-10-140-472-10	Sequence 10, Appl
12	37.6	3.3	594	US-10-141-761-10	Sequence 10, Appl
13	37.6	3.3	594	US-10-142-885-10	Sequence 10, Appl
14	37.6	3.3	594	US-10-158-790-10	Sequence 10, Appl
15	37.6	3.3	594	US-10-123-155-10	Sequence 10, Appl
16	37.6	3.3	594	US-10-146-731-10	Sequence 10, Appl

17	37.4	3.3	4447	9	US-09-835-381-5	Sequence 5, Appli
18	37.4	3.3	4447	9	US-09-835-381-7	Sequence 7, Appli
19	37.4	3.3	4447	14	US-10-196-232-5	Sequence 5, Appli
20	37.4	3.3	4447	14	US-10-196-232-6	Sequence 6, Appli
21	36	3.2	299	14	US-10-184-644-164	Sequence 164, App
22	36	3.2	299	14	US-10-184-634-164	Sequence 164, App
23	34.8	3.1	1631	9	US-09-912-020-208	Sequence 208, App
24	34.4	3.0	999	14	US-10-184-644-434	Sequence 434, App
25	34.4	3.0	999	14	US-10-184-634-434	Sequence 434, App
26	34.2	3.0	623	12	US-10-027-632-11021	Sequence 11021, A
27	34.2	3.0	623	13	US-10-027-632-11021	Sequence 11021, A
28	34	3.0	1049	12	US-10-140-472-358	Sequence 358, App
29	34	3.0	1049	12	US-10-141-761-358	Sequence 358, App
30	34	3.0	1049	12	US-10-142-885-358	Sequence 358, App
31	34	3.0	1049	12	US-10-158-790-358	Sequence 358, App
32	34	3.0	1049	14	US-10-123-155-358	Sequence 358, App
33	34	3.0	1049	15	US-10-146-731-358	Sequence 358, App
34	33.6	2.9	1134	13	US-10-021-788-3	Sequence 3, Appli
35	33.6	2.9	1338	10	US-09-738-626-1200	Sequence 1200, Ap
36	33.6	2.9	1972	12	US-10-017-161-1981	Sequence 1981, Ap
37	33.6	2.9	3309400	10	US-09-738-626-1	Sequence 1, Appli
38	33.6	2.9	9025608	14	US-10-017-161-1781	Sequence 1781, Ap
39	33.4	2.9	1390	12	US-09-901-152-3	Sequence 3, Appli
40	33.4	2.9	58985	11	US-09-855-824-3	Sequence 358, App
41	33.4	2.9	143601	12	US-09-991-936-1833	Sequence 1833, Ap
42	33.2	2.9	549	11	US-10-349-806-3	Sequence 3, Appli
43	33.2	2.9	1798	11	US-09-938-795A-4	Sequence 4, Appli
44	33.2	2.9	1798	11	US-09-946-374-83	Sequence 83, Appli
45	33.2	2.9	1844	11		

## ALIGNMENTS

RESULT 1  
US-10-007-527A-1  
; Sequence 1, Application US/10007527A  
; Publication No. US20030044807A1  
; GENERAL INFORMATION:  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: C14709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,527A  
; PRIOR FILING DATE: 2001-12-05  
; CURRENT APPLICATION NUMBER: 60/254,868  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 1  
; LENGTH: 1140  
; TYPE: DNA  
; ORGANISM: Rhodococcus AM12  
US-10-007-527A-1

Query Match	Best Local Similarity	100.0%	Pred. No.	Score	1140;	DB	14;	Length	1140;
Matches	1140;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	ATGACGAGCGTAAGTCGTAACACCTTTCGGCAAGACCGGCTCCGTCGTGCG	60						
DB	1	ATGACGAGCGTAAGTCGTAACACCTTTCGGCAAGACCGGCTCCGTCGTGCG	60						
QY	61	TCGGATTAACGCGGCGCATCCGCGACGACCTGCAACCACTTCAACCAATCCACGTC	120						
DB	61	TCGGATTAACGCGGCGCATCCGCGACGACCTGCAACCACTTCAACCAATCCACGTC	120						
QY	121	GAAACATTTAAAGCCCTGTGCGCGCGCGATTTCTGGGCGTAAGCGTGTGCAATGCAAC	180						
DB	121	GAAACATTTAAAGCCCTGTGCGCGCGCGATTTCTGGGCGTAAGCGTGTGCAATGCAAC	180						

QY 181 GGTCCGAAAGTTCTGGATTCCGAGGCTTCTGTTCTCTGCGAAAGGGCTGATCTGCCCC 240  
DB 181 GGTCCGAAAGTTCTGGATTCCGAGGCTTCTGTTCTCTGCGAAAGGGCTGATCTGCCCC 240  
QY 241 TGTGTGCGGGAAGAAATCGGTGACATCGTGCAGAGAAATTTCTAACTGTGTGCTCAT 300  
DB 241 TGTGTGCGGGAAGAAATCGGTGACATCGTGCAGAGAAATTTCTAACTGTGTGCTCAT 300  
QY 301 CAACCTGGGACTGATCTGTGGATGTGACAGATGACATGACCATGCGCATCAAGTGTGTGAC 360  
DB 301 CAACCTGGGACTGATCTGTGGATGTGACAGATGACATGACCATGCGCATCAAGTGTGTGAC 360  
QY 361 CGGCTCCACGACCTATGACCTGACCTTTGCGAGCCTTGGAAAGCTGCAACCAAGTGTGTG 420  
DB 361 CGGCTCCACGACCTATGACCTGACCTTTGCGAGCCTTGGAAAGCTGCAACCAAGTGTGTG 420  
QY 421 CGTGGCGGTACGGAAGGTGAATGTACGCTGTGACAGATGACATGACCATGCGCATCAAGT 480  
DB 421 CGTGGCGGTACGGAAGGTGAATGTACGCTGTGACAGATGACATGACCATGCGCATCAAGT 480  
QY 481 ACTCAGGAAAAAGGCTGACGCTGACGCTTCAAGCTTCAAGCTGCTCACTCATGTTCAAGTGTGAC 540  
DB 481 ACTCAGGAAAAAGGCTGACGCTGACGCTTCAAGCTTCAAGCTGCTCACTCATGTTCAAGTGTGAC 540  
QY 541 GTGAGTGAAGAACCTCTGCAATCTTCTGCGATGCGATGCTGATGCTGAGCTTCCAA 600  
DB 541 GTGAGTGAAGAACCTCTGCAATCTTCTGCGATGCGATGCTGATGCTGAGCTTCCAA 600  
QY 601 CTGCTATCTCTGGGATTTGCTGCGCAGATGCTGATGCTGATGCTGATGCTGATGCTGATG 660  
DB 601 CTGCTATCTCTGGGATTTGCTGCGCAGATGCTGATGCTGATGCTGATGCTGATGCTGATG 660  
QY 661 ATCGGCGGTGAAGCTGATCAAGTCTCGCTGCTATCTGACGAAATTCATCTGAGCTT 720  
DB 661 ATCGGCGGTGAAGCTGATCAAGTCTCGCTGCTATCTGACGAAATTCATCTGAGCTT 720  
QY 721 GGTATGAGAGTTGTGTGATGCGACGAGAAAAAGTGTGACATGCAACCGTGTGACCCCTG 780  
DB 721 GGTATGAGAGTTGTGTGATGCGACGAGAAAAAGTGTGACATGCAACCGTGTGACCCCTG 780  
QY 781 GAAATGCTGTTTATGATGAGTGGGCGGAGATCCCAAGCGTGTGACATGCTGTGCGAGAAATTT 840  
DB 781 GAAATGCTGTTTATGATGAGTGGGCGGAGATCCCAAGCGTGTGACATGCTGTGCGAGAAATTT 840  
QY 841 GAGTTGTTGATGATGAGGAGCTGCGGCAATGCGTGTGCTGCGTGTGCTGCTGCTGCTGCT 900  
DB 841 GAGTTGTTGATGATGAGGAGCTGCGGCAATGCGTGTGCTGCGTGTGCTGCTGCTGCTGCT 900  
QY 901 GGTCTTGGGCGAGAACTTAACAGATGCTCAAGTCTGTTGACGAGAGAAATCTGCGCGCT 960  
DB 901 GGTCTTGGGCGAGAACTTAACAGATGCTCAAGTCTGTTGACGAGAGAAATCTGCGCGCT 960  
QY 961 ATGCTTGGGCTGATCTTCTGCGGCGATGCTGATGATGATGATGATGATGATGATGATGATG 1020  
DB 961 ATGCTTGGGCTGATCTTCTGCGGCGATGCTGATGATGATGATGATGATGATGATGATGATG 1020  
QY 1021 TTGCGGAGAGATCTCTGCGACTGTCGAAAGTGGCGGCACTTGGGAAATCTTCTGATGAC 1080  
DB 1021 TTGCGGAGAGATCTCTGCGACTGTCGAAAGTGGCGGCACTTGGGAAATCTTCTGATGAC 1080  
QY 1081 TTGCAATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
DB 1081 TTGCAATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140

## RESULT 2

US-10-007-452-1  
; Sequence 1, Application US/10007452  
; Publication No. US20030093701A1  
; GENERAL INFORMATION:  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong

; APPLICANT: Kostichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: CL1709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,452  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 1  
; LENGTH: 1140  
; TYPE: DNA  
; ORGANISM: Rhodococcus AM12  
US-10-007-452-1

Query Match 100.0%; Score 1140; DB 14; Length 1140;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACCAAGCGTAAGTGTGAAACCTTTCCGCAAGAACCGGCTCCGCTCGTGTGCG 60  
DB 1 ATGACCAAGCGTAAGTGTGAAACCTTTCCGCAAGAACCGGCTCCGCTCGTGTGCG 60  
QY 61 TCCGATTAAGCGCGCATCCGCGACGATCTGCGACCCCAACTTCAACAAATCCACGTCA 120  
DB 61 TCCGATTAAGCGCGCATCCGCGACGATCTGCGACCCCAACTTCAACAAATCCACGTCA 120  
QY 121 GAAACATTTAAAGCGCTGTGAGCGCGCGCATTTCTGCGGTGAACGAGTGTGACATGTCAC 180  
DB 121 GAAACATTTAAAGCGCTGTGAGCGCGCGCATTTCTGCGGTGAACGAGTGTGACATGTCAC 180  
QY 181 GGTCCGAAAGTTCTGGAATTCGAGGCTTCTGTTCTGCGAAAGGGCTGATCTGCCCC 240  
DB 181 GGTCCGAAAGTTCTGGAATTCGAGGCTTCTGTTCTGCGAAAGGGCTGATCTGCCCC 240  
QY 241 TGTGTGCGGGAAGAAATCGGTGACATCGTGCAGAGAAATTTCTAACTGTGTGCTCAT 300  
DB 241 TGTGTGCGGGAAGAAATCGGTGACATCGTGCAGAGAAATTTCTAACTGTGTGCTCAT 300  
QY 301 CAACCTGGGACTGATCTGTGGATGTGACAGATGACATGACCATGCGCATCAAGTGTGTGAC 360  
DB 301 CAACCTGGGACTGATCTGTGGATGTGACAGATGACATGACCATGCGCATCAAGTGTGTGAC 360  
QY 361 CGGCTCCACGACCTATGACCTTGGAGCCTTGGAAAGCTGCAACCAAGTGTGTG 420  
DB 361 CGGCTCCACGACCTATGACCTTGGAGCCTTGGAAAGCTGCAACCAAGTGTGTG 420  
QY 421 CGTGGCGGTACGGAAGGTGAATGTACGCTGTGACAGATGACATGACCATGCGCATCAAGT 480  
DB 421 CGTGGCGGTACGGAAGGTGAATGTACGCTGTGACAGATGACATGACCATGCGCATCAAGT 480  
QY 481 ACTCAGGAAAAAGGCTGACGCTGACGCTTCAAGCTTCAAGCTGCTCACTCATGTTCAAGTGTGAC 540  
DB 481 ACTCAGGAAAAAGGCTGACGCTGACGCTTCAAGCTTCAAGCTGCTCACTCATGTTCAAGTGTGAC 540  
QY 541 GTGAGTGAAGAACCTCTGCAATCTTCTGCGATGCGATGCTGATGCTGAGCTTCCAA 600  
DB 541 GTGAGTGAAGAACCTCTGCAATCTTCTGCGATGCGATGCTGATGCTGAGCTTCCAA 600  
QY 601 CTGCTATCTCTGGGATTTGCTGCGCAGATGCTGATGCTGATGCTGATGCTGATGCTGATG 660  
DB 601 CTGCTATCTCTGGGATTTGCTGCGCAGATGCTGATGCTGATGCTGATGCTGATGCTGATG 660  
QY 661 ATCGGCGGTGAAGCTGATCAAGTCTCGCTGCTATCTGACGAAATTCATCTGAGCTT 720  
DB 661 ATCGGCGGTGAAGCTGATCAAGTCTCGCTGCTATCTGACGAAATTCATCTGAGCTT 720  
QY 721 GGTATGAGAGTTGTGTGATGCGACGAGAAAAAGTGTGACATGCAACCGTGTGACCCCTG 780  
DB 721 GGTATGAGAGTTGTGTGATGCGACGAGAAAAAGTGTGACATGCAACCGTGTGACCCCTG 780  
QY 781 GAAATGCTGTTTATGATGAGTGGGCGGAGATCCCAAGCGTGTGACATGCTGTGCGAGAAATTT 840

Db 781 GAATCGCTGTGATGAGTGGGGGAGATCCAAAGCGTTGGAACTGTGGCAGAAATTT 840  
QY 841 GAGTTTGGTTGATGAGGAGCTGGGCAATGCGTGTGCCGTGATTTGCCGCCAGCT 900  
Db 841 GAGTTTGGTTGATGAGGAGCTGGGCAATGCGTGTGCCGTGATTTGCCGCCAGCT 900  
QY 901 GGTCTTGGGGCAGAACTTAAGATGCTCAGATGTTGAGAGAGAAATTTGCCCGGTC 960  
Db 901 GGTCTTGGGGCAGAACTTAAGATGCTCAGATGTTGAGAGAGAAATTTGCCCGGTC 960  
QY 961 ATGTTGGCGATCATTCGCGGCGCATGATGATGATTTGCACTTGTGGCCTTACGTC 1020  
Db 961 ATGTTGGCGATCATTCGCGGCGCATGATGATGATTTGCACTTGTGGCCTTACGTC 1020  
QY 1021 TTGCGGAGATCTCTCGGACTCTGTCGAAAGCTGGCGCACTTGGGAAAACTTCTGTGATCA 1080  
Db 1021 TTGCGGAGATCTCTCGGACTCTGTCGAAAGCTGGCGCACTTGGGAAAACTTCTGTGATCA 1080  
QY 1081 TTGCATTTATGATTTGCCCGCAGCGATGTGGGCCCCCGATTAATATGCTTGGCAAGTGA 1140  
Db 1081 TTGCATTTATGATTTGCCCGCAGCGATGTGGGCCCCCGATTAATATGCTTGGCAAGTGA 1140

## RESULT 3

US-10-007-527A-5/c  
; Sequence 5, Application US/10007527A  
; Publication No. US20030044807A1  
; GENERAL INFORMATION:  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: Kostichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: C11709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,527A  
; PRIOR FILING DATE: 2001-12-05  
; PRIOR APPLICATION NUMBER: 60/254,868  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 5  
; LENGTH: 6334  
; TYPE: DNA  
; ORGANISM: Rhodococcus AN12  
US-10-007-527A-5

Query Match 100.0%; Score 1140; DB 14; Length 6334;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGACGCTAAGTGTGGAACACCTTCCGCGCAAAAGACGCGCTCCCGTCTGTGTG 60  
Db 3051 ATGACGACGCTAAGTGTGGAACACCTTCCGCGCAAAAGACGCGCTCCCGTCTGTGTG 2992  
QY 61 TCCGATTAAGCGCGCATCCGCGCAGAACTGGGACCCAACTTCAACAATCACCAGTCA 120  
Db 2991 TCCGATTAAGCGCGCATCCGCGCAGAACTGGGACCCAACTTCAACAATCACCAGTCA 2932  
QY 121 GAAACATTTAACGCTGTGGCGCGGCGGATTTCTGGCGTGAACGCTGTGACCATTTGTAAC 180  
Db 2931 GAAACATTTAACGCTGTGGCGCGGCGGATTTCTGGCGTGAACGCTGTGACCATTTGTAAC 2872  
QY 181 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTGCTTCTGCGGAAAGGCTGATTCGCC 240  
Db 2871 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTGCTTCTGCGGAAAGGCTGATTCGCC 2812  
QY 241 TGCTGTGCGGAAAAAGTGTGCAATCGTGCAGACGAAATTTTCAAGTTGTGTCAT 300  
Db 2811 TGCTGTGCGGAAAAAGTGTGCAATCGTGCAGACGAAATTTTCAAGTTGTGTCAT 2752  
QY 301 CAACTCGGGACTGATCTGTGCGATGTGACATGACCAATGCCCATACAGCTGTGTAG 360  
Db 2751 CAACTCGGGACTGATCTGTGCGATGTGACATGACCAATGCCCATACAGCTGTGTAG 2692

QY 361 CGGCTTCACAGACCTTATGACTGAGACTTTTGGGAGGCTTGAAGCTGCCACCAAGGTCGT 420  
Db 2691 CGGCTTCACAGACCTTATGACTGAGACTTTTGGGAGGCTTGAAGCTGCCACCAAGGTCGT 2632  
QY 421 CGTTGGCGTACGGAACGTGAAATGTAACGCTCGGACGGAATACGTCGCGCTGTGAAATC 480  
Db 2631 CGTTGGCGTACGGAACGTGAAATGTAACGCTCGGACGGAATACGTCGCGCTGTGAAATC 2572  
QY 481 ACTCAAGAAAAACGCTGTGACAGCTTCACGCTTCAACGCTTCACTGTTCAAGTGTGAC 540  
Db 2571 ACTCAAGAAAAACGCTGTGACAGCTTCACGCTTCAACGCTTCACTGTTCAAGTGTGAC 2512  
QY 541 GTGAGTGAAGAACCTCTGCAATCTTCTGTGATGCAATGTTGATGTGTGACTTCAAA 600  
Db 2511 GTGAGTGAAGAACCTCTGCAATCTTCTGTGATGCAATGTTGATGTGTGACTTCAAA 2452  
QY 601 CTGCTATCTCTGGGATTTGCTGCGGACCACTACGAAATTTGGGAGTGTGATGTAAGAA 660  
Db 2451 CTGCTATCTCTGGGATTTGCTGCGGACCACTACGAAATTTGGGAGTGTGATGTAAGAA 2392  
QY 661 ATCGGCGTGAAGCTGATCAAGTTCTGCTGCTATCTGACGAAATTTGATCTGGCGTT 720  
Db 2391 ATCGGCGTGAAGCTGATCAAGTTCTGCTGCTATCTGACGAAATTTGATCTGGCGTT 2332  
QY 721 GGTATGAGAGTTGTGATGTGCGGACGGAAGAAAGTGTGCAATGCGCAACCGTGTG 780  
Db 2331 GGTATGAGAGTTGTGATGTGCGGACGGAAGAAAGTGTGCAATGCGCAACCGTGTG 2272  
QY 781 GAAATCGCTGTGATGTCAGTGGGCGGGGATCCACAGCGTTGGAACCTGTGGCAGAAAT 840  
Db 2271 GAAATCGCTGTGATGTCAGTGGGCGGGGATCCACAGCGTTGGAACCTGTGGCAGAAAT 2212  
QY 841 GAGTTTGTTCATGATGAGAGCTCGGCAATCGCTGTGCTCCGATGTTGCCGTGCCAGCT 900  
Db 2211 GAGTTTGTTCATGATGAGAGCTCGGCAATCGCTGTGCTCCGATGTTGCCGTGCCAGCT 2152  
QY 901 GGTCTTGGGCGAAGCTTAACAGATGCTGAGATGTTGAGAGAGAAATTTGCCCGGTC 960  
Db 2151 GGTCTTGGGCGAAGCTTAACAGATGCTGAGATGTTGAGAGAGAAATTTGCCCGGTC 2092  
QY 961 ATGTTGGCATTCATTCGCGGCGCATGCTGTGATGATGTTGCACTTGTGGCCTTACGTC 1020  
Db 2091 ATGTTGGCATTCATTCGCGGCGCATGCTGTGATGATGTTGCACTTGTGGCCTTACGTC 2032  
QY 1021 TTGCGGAGATCTCTCGGACTCTGTCGAAAGCTGGCGCACTTGGGAAAACTTCTGTGATCA 1080  
Db 2031 TTGCGGAGATCTCTCGGACTCTGTCGAAAGCTGGCGCACTTGGGAAAACTTCTGTGATCA 1972  
QY 1081 TTGCATTTATGATTTGCCCGCAGCGATGTGGGCCCCCGATTAATATGCTTGGCAAGTGA 1140  
Db 1971 TTGCATTTATGATTTGCCCGCAGCGATGTGGGCCCCCGATTAATATGCTTGGCAAGTGA 1912

## RESULT 4

US-10-007-452-5/c  
; Sequence 5, Application US/10007452  
; Publication No. US20030093701A1  
; GENERAL INFORMATION:  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: Kostichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: C11709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,452  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: 60/254,868  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 5  
; LENGTH: 6334





Db 4572 ACTCAGGAAAAACGGCTGAGCAGTCCAGCTTCAAGCGCTACTCATGTTTCAAGTGAGC 4513  
QY 541 GTGAGTGAGAACATCTCTGAAATCTTCTGGAATGCGAATGTTGATCGGTGGAATTCCAA 600  
Db 4512 GTGAGTGAGAACATCTCTGAAATCTTCTGGAATGCGAATGTTGATCGGTGGAATTCCAA 4453  
QY 601 CTCGTATCTCTGGAATTTTGTGCGGCACTACGTAATTCGGAGTGTCTCGATGTAAGAAAG 660  
Db 4452 CTCGTATCTCTGGAATTTTGTGCGGCACTACGTAATTCGGAGTGTCTCGATGTAAGAAAG 4393  
QY 661 ATCGGCGGTGAAGCTGATCAAGTCTCTGCTGATCTGACGAAATTTGATCTGAGCTT 720  
Db 4392 ATCGGCGGTGAAGCTGATCAAGTCTCTGCTGATCTGACGAAATTTGATCTGAGCTT 4333  
QY 721 GGTATGAGAGTTGTTAGTGGCGGACGGAATAAGTGTGACATGCGCAACCGTGCACCTG 780  
Db 4332 GGTATGAGAGTTGTTAGTGGCGGACGGAATAAGTGTGACATGCGCAACCGTGCACCTG 4273  
QY 781 GAAATCGCTGTGATGAGCAGTGGCGGAGATCCACAAAGCGTTGAACTGTGGCGAAATTT 840  
Db 4272 GAAATCGCTGTGATGAGCAGTGGCGGAGATCCACAAAGCGTTGAACTGTGGCGAAATTT 4213  
QY 841 GAGTTTGTGATGAGAGTGGAGTGGGCAATCGCTGTGCTCCGTGATTTGCCAGCT 900  
Db 4212 GAGTTTGTGATGAGAGTGGAGTGGGCAATCGCTGTGCTCCGTGATTTGCCAGCT 4153  
QY 901 GGTCTTTGGGCGAACTTAACAGATGCTGATGTTGAGCAGAAAGATTTGCCCGCT 960  
Db 4152 GGTCTTTGGGCGAACTTAACAGATGCTGATGTTGAGCAGAAAGATTTGCCCGCT 4093  
QY 961 ATGTTGCGATCATTCGGCGGCGGATCGTGGATGATGATTCGACTTGGCGCTTAACGTC 1020  
Db 4092 ATGTTGCGATCATTCGGCGGCGGATCGTGGATGATGATTCGACTTGGCGCTTAACGTC 4033  
QY 1021 TTCCGCGAGATCTCTGGAATCTGTCGAAGCTGGCGGACTTTGGGAAATCTTGTGATCAC 1080  
Db 4032 TTCCGCGAGATCTCTGGAATCTGTCGAAGCTGGCGGACTTTGGGAAATCTTGTGATCAC 3973  
QY 1081 TTGCATTATCGATTGCCCGGCGAGCGGATGTGCGCCCGGATTAATTCGTTTGCAGTGA 1140  
Db 3972 TTGCATTATCGATTGCCCGGCGAGCGGATGTGCGCCCGGATTAATTCGTTTGCAGTGA 3913

RESULT 6  
US-10-007-452-7/c  
; Sequence 7, Application US/10007452  
; Publication No. US20030093701A1  
; GENERAL INFORMATION:  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Bramucci, Michael G.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: Koestlichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: CL1709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,452  
; PRIOR APPLICATION NUMBER: 60/254,868  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 7  
; LENGTH: 9652  
; TYPE: DNA  
; ORGANISM: P1aamid prHR17  
US-10-007-452-7

Query Match 100.0%; Score 1140; DB 14; Length 9652;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGCTGATGCTGACACCTTCCGCGAAAGACGCGCTCCGCTCGTGTG 60  
Db 5052 ATGACGAGCTGATGCTGACACCTTCCGCGAAAGACGCGCTCCGCTCGTGTG 4993

QY 61 TCCGATTAACCGGCGATCCGCGACGAACTGCGAACCCAACTTCAACAAATCCACAGTCA 120  
Db 4992 TCCGATTAACCGGCGATCCGCGACGAACTGCGAACCCAACTTCAACAAATCCACAGTCA 4933  
QY 121 GAAACATTTAAACGCTGGGCGCGCGGATTTCTGGCGTGAACGCTGTGACATTTGTCAAC 180  
Db 4932 GAAACATTTAAACGCTGGGCGCGCGGATTTCTGGCGTGAACGCTGTGACATTTGTCAAC 4873  
QY 181 GGTCCGAAAGGTTCTGATTCGAGGCGCTTGCTTCTGCGGAAAGGCGTGAATCTGCCCC 240  
Db 4872 GGTCCGAAAGGTTCTGATTCGAGGCGCTTGCTTCTGCGGAAAGGCGTGAATCTGCCCC 4813  
QY 241 TGTCTGCGGAAAAAGTGGTGCACATGCTGCGAGACGAAATTTCTCAAGTTTGTCTCAT 300  
Db 4812 TGTCTGCGGAAAAAGTGGTGCACATGCTGCGAGACGAAATTTCTCAAGTTTGTCTCAT 4753  
QY 301 CAATCGGAGCTGATCTGTTGCGATGGTGAACATGACCATGCGCATACAGCTGTGAC 360  
Db 4752 CAATCGGAGCTGATCTGTTGCGATGGTGAACATGACCATGCGCATACAGCTGTGAC 4693  
QY 361 CGGCTCCAGACCTATGAGCTGACTTTCGGGAGCGTGGAAAGCTGCGACCAACGCTGT 420  
Db 4692 CGGCTCCAGACCTATGAGCTGACTTTCGGGAGCGTGGAAAGCTGCGACCAACGCTGT 4633  
QY 421 GGTGGCGTACGGAACGTGAATGTACGCTGCGACGATACGTCGCGCTGTTGAAATC 480  
Db 4632 GGTGGCGTACGGAACGTGAATGTACGCTGCGACGATACGTCGCGCTGTTGAAATC 4573  
QY 481 ACTCAGGAAAAACGGCTGGGACGTCACGTTCAACGCTTCAATGTTCAAGTGTGAC 540  
Db 4572 ACTCAGGAAAAACGGCTGGGACGTCACGTTCAACGCTTCAATGTTCAAGTGTGAC 4513  
QY 541 GTGAGTGAGAACATCTCTGAAATCTTCTCGAATGCGAATGTTGATCGGTGGAATTCCAA 600  
Db 4512 GTGAGTGAGAACATCTCTGAAATCTTCTCGAATGCGAATGTTGATCGGTGGAATTCCAA 4453  
QY 601 CTCGTATCTCTGGAATTTTGTGCGGCACTACGTAATTCGGAGTGTCTCGATGTAAGAAAG 660  
Db 4452 CTCGTATCTCTGGAATTTTGTGCGGCACTACGTAATTCGGAGTGTCTCGATGTAAGAAAG 4393  
QY 661 ATCGGCGGTGAAGCTGATCAAGTCTCTGCTGATCTGACGAAATTTGATCTGAGCTT 720  
Db 4392 ATCGGCGGTGAAGCTGATCAAGTCTCTGCTGATCTGACGAAATTTGATCTGAGCTT 4333  
QY 721 GGTATGAGAGTTGTTAGTGGCGGACGGAATAAGTGTGACATGCGCAACCGTGCACCTG 780  
Db 4332 GGTATGAGAGTTGTTAGTGGCGGACGGAATAAGTGTGACATGCGCAACCGTGCACCTG 4273  
QY 781 GAAATCGCTGTGATGAGCAGTGGCGGAGATCCACAAAGCGTTGAACTGTGGCGAAATTT 840  
Db 4272 GAAATCGCTGTGATGAGCAGTGGCGGAGATCCACAAAGCGTTGAACTGTGGCGAAATTT 4213  
QY 841 GAGTTTGTGATGAGAGTGGAGTGGGCAATCGCTGTGCTCCGTGATTTGCCAGCT 900  
Db 4212 GAGTTTGTGATGAGAGTGGAGTGGGCAATCGCTGTGCTCCGTGATTTGCCAGCT 4153  
QY 901 GGTCTTTGGGCGAACTTAACAGATGCTGATGTTGAGCAGAAAGATTTGCCCGCT 960  
Db 4152 GGTCTTTGGGCGAACTTAACAGATGCTGATGTTGAGCAGAAAGATTTGCCCGCT 4093  
QY 961 ATGTTGCGATCATTCGGCGGCGGATCGTGGATGATGATTCGACTTGGCGCTTAACGTC 1020  
Db 4092 ATGTTGCGATCATTCGGCGGCGGATCGTGGATGATGATTCGACTTGGCGCTTAACGTC 4033  
QY 1021 TTCCGCGAGATCTCTGGAATCTGTCGAAGCTGGCGGACTTTGGGAAATCTTGTGATCAC 1080  
Db 4032 TTCCGCGAGATCTCTGGAATCTGTCGAAGCTGGCGGACTTTGGGAAATCTTGTGATCAC 3973  
QY 1081 TTGCATTATCGATTGCCCGGCGAGCGGATGTGCGCCCGGATTAATTCGTTTGCAGTGA 1140  
Db 3972 TTGCATTATCGATTGCCCGGCGAGCGGATGTGCGCCCGGATTAATTCGTTTGCAGTGA 3913

```
RESULT 7
US-10-007-527a-6/c
; Sequence 6, Application US/10007527a
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527a
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 11241
; TYPE: DNA
; ORGANISM: Plasmid pRHR17
US-10-007-527a-6
```

```
Query Match      100.0%; Score 1140; DB 14; Length 11241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGACGAGGTAAAGTCTGACACCTTCCGGCAAGACGGGCTCCGCTCGTGTG 60
DB 6641 ATGACGAGGTAAAGTCTGACACCTTCCGGCAAGACGGGCTCCGCTCGTGTG 6582
QY 61 TCCGATTAAGCGCGGCAATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAAGTCA 120
DB 6581 TCCGATTAAGCGCGGCAATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAAGTCA 6522
QY 121 GAAACATTTAAGCCTGTGCGCGCGCGGCTTCTGCGGTGAACGCTGTGCAAC 180
DB 6521 GAAACATTTAAGCCTGTGCGCGCGGCTTCTGCGGTGAACGCTGTGCAAC 6462
QY 181 GGTCCGAAAGTTCGATTCGAGAGGCTTCTGCTGGGAAAGGGCTGATCTGCCCC 240
DB 6461 GGTCCGAAAGTTCGATTCGAGAGGCTTCTGCTGGGAAAGGGCTGATCTGCCCC 6402
QY 241 TCGTGTGCGGAAAGTTCGATTCGAGAGGCTTCTGCTGGGAAAGGGCTGATCTGCCCC 300
DB 6401 TCGTGTGCGGAAAGTTCGATTCGAGAGGCTTCTGCTGGGAAAGGGCTGATCTGCCCC 6342
QY 301 CAACCTGGGACTGATCTGCTGGGATGTGACGATGACCAATGCGGCATACAGCTGTGAG 360
DB 6341 CAACCTGGGACTGATCTGCTGGGATGTGACGATGACCAATGCGGCATACAGCTGTGAG 6282
QY 361 CGGCTCCAGACCTATGACATGACCTTCCGGCAAGACGGGCTCCGCTCGTGTG 420
DB 6281 CGGCTCCAGACCTATGACATGACCTTCCGGCAAGACGGGCTCCGCTCGTGTG 6222
QY 421 CGTGGCGTACGGAAGTGAATGTACGCGCTGCGACGATACGCTGTGGAATC 480
DB 6221 CGTGGCGTACGGAAGTGAATGTACGCGCTGCGACGATACGCTGTGGAATC 6162
QY 481 ACTCAGGAAAAAAGCGCTGCGACGCTCAAGTTCAGGCTACTCATGTTCAGTGTGAC 540
DB 6161 ACTCAGGAAAAAAGCGCTGCGACGCTCAAGTTCAGGCTACTCATGTTCAGTGTGAC 6102
QY 541 GTGAGTGAAGACATCTCTGCAATCTCTGCGATGCGATGATGATGCGGTGCAAA 600
DB 6101 GTGAGTGAAGACATCTCTGCAATCTCTGCGATGCGATGATGATGCGGTGCAAA 6042
QY 601 CTGATCTCTGGAATTTGCTGCGCACTACGTAATTCGAGTGTCTCATGTGACGAAG 660
DB 6041 CTGATCTCTGGAATTTGCTGCGCACTACGTAATTCGAGTGTCTCATGTGACGAAG 5982
QY 661 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGACGAAAAATTGATCTGCGCTT 720
```

```
DB 5981 ATCGGCGGTGAAGCTGATCAAGTTCTGCTGCGTATCTGACGAAAAATTGATCTGCGCTT 5922
QY 721 GGTATGAGGTGATGATGCGGACGCAAGAAAGGTGATGACATGCGCAACCGTGCACCTTGG 780
DB 5921 GGTATGAGGTGATGATGCGGACGCAAGAAAGGTGATGACATGCGCAACCGTGCACCTTGG 5862
QY 781 GAAATCGCTGTGATGACAGTGGGCGGAGATCAAGGCTGTGAATCTGTGCGAGATTT 840
DB 5861 GAAATCGCTGTGATGACAGTGGGCGGAGATCAAGGCTGTGAATCTGTGCGAGATTT 5802
QY 841 GAGTTGGTGTGATGAGGACGCTGCGCAATTCGCTGTGCTCCGCTGATTCGCGCCGAGCT 900
DB 5801 GAGTTGGTGTGATGAGGACGCTGCGCAATTCGCTGTGCTCCGCTGATTCGCGCCGAGCT 5742
QY 901 GGTCTGGGCGGACAACTAAGTCTGAGATGCTGAGAGCAAGAAATTCGCGCGCTC 960
DB 5741 GGTCTGGGCGGACAACTAAGTCTGAGATGCTGAGAGCAAGAAATTCGCGCGCTC 5682
QY 961 ATGATTTGATCATTTCCGCGCGGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 5681 ATGATTTGATCATTTCCGCGCGGATCTGATGATGATGATGATGATGATGATGATGATGATGAT 5622
QY 1021 TTGCGGAGATCTTCCGCGGACGCTGCGCAATTCGCTGTGCTCCGCTGATTCGCGCCGAGCT 1080
DB 5621 TTGCGGAGATCTTCCGCGGACGCTGCGCAATTCGCTGTGCTCCGCTGATTCGCGCCGAGCT 5562
QY 1081 TTGCAATTAAGTTCGCGGACGCGGATGATGCGGCGCGGATTAATTCGCTGTGCAAGTGA 1140
DB 5561 TTGCAATTAAGTTCGCGGACGCGGATGATGCGGCGCGGATTAATTCGCTGTGCAAGTGA 5502
```

```
RESULT 8
US-10-007-452-6/c
; Sequence 6, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 11241
; TYPE: DNA
; ORGANISM: Plasmid pRHR17
US-10-007-452-6
```

```
Query Match      100.0%; Score 1140; DB 14; Length 11241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGACGAGGTAAAGTCTGACACCTTCCGGCAAGACGGGCTCCGCTCGTGTG 60
DB 6641 ATGACGAGGTAAAGTCTGACACCTTCCGGCAAGACGGGCTCCGCTCGTGTG 6582
QY 61 TCCGATTAAGCGCGGCAATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAAGTCA 120
DB 6581 TCCGATTAAGCGCGGCAATCCGCGACGAACTGCGACCCAACTTCAACAAATCAACAAGTCA 6522
QY 121 GAAACATTTAAGCCTGTGCGCGCGCGGCTTCTGCGGTGAACGCTGTGCAAC 180
DB 6521 GAAACATTTAAGCCTGTGCGCGCGCGGCTTCTGCGGTGAACGCTGTGCAAC 6462
QY 181 GGTCCGAAAGTTCGATTCGAGAGGCTTCTGCTGGGAAAGGGCTGATCTGCCCC 240
```

```

1  APPLICANT: Gurney, Austin J.
2  APPLICANT: Pan, James
3  APPLICANT: Smith, Victoria
4  APPLICANT: Watanabe, Colin K.
5  APPLICANT: Wood, William I.
6  APPLICANT: Zhang, Zemin
7  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
8  TITLE OF INVENTION: ACIDS ENCODING THE SAME
9  FILE REFERENCE: P3430R1C27
10 CURRENT APPLICATION NUMBER: US/10/184,644
11 PRIOR APPLICATION removed - See File Wrapper or Palm
12 NUMBER OF SEQ ID NOS: 612
13 SEQ ID NO 332
14 LENGTH: 520
15 TYPE: PRT
16 ORGANISM: Homo Sapien
17 US-10-184-644-332
18
19 Query Match      3.6%  Score 40.8; DB 14; Length 520;
20 Best Local Similarity 18.5%; Pred. No. 0.0098;
21 Matches 74; Conservative 81; Mismatches 246; Indels 0; Gaps 0
22
23 QY 522 ACTCATGTTTCAGTGTGACGTGAGGAGAAATCCTCCAAATCCTTCGTGATGATGTT 581
24   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
25 DB DSPSPSLIQSHPRHHLLAQASRLQVLAQLTWKVRVSHHLLQRVNDTQNPMMRRIRGE 149
26
27 QY 582 CGATCGGTGACCTTCCAAACTGTATCTCTGGGATTTCTGCGCACACTAGCTAATTTCGG 641
28   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
29 DB 150 QGAAGLQGHKKAMMPGAPPPGPPEAEKAGKGMGRDGTGSPGQPPVGEAGLQGP 209
30
31 QY 642 TGCTGTGATGATGAGAAAGATCGGCGGTGAAGCTGATCAAGTTCCTGCTGCTATCTGAC 701
32   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
33 DB 210 QGAAGKQATGTPPQGEKSKGQGLIGPYGETGTEKEKPDGLGPSKGRGMGDAV 269
34
35 QY 702 GAAATTCGATCTGCGCTTGTGTATGAGGTTGTAGTGGCAGCGAAAGAGTGTGAC 761
36   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
37 DB 270 MGPPGAGQSKGDFRPPRPPGLAGRPAAKGDQGGELQGVPPPAVGHGAKGEPSGSG 329
38
39 QY 762 TGGCAACCGTGCACCCCTGGGAAATCGCTGTGTATGACAGTGGCGGGGATCCACAAGCGTT 821
40   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
41 DB 330 PGRAGLPGSPGSPGATGLKSGKDPDGLQGGGRKGSVGPAPAVKGEGSPGLAGPKGA 389
42
43 QY 822 GGAACGTGGCGGAATTTGAGTTTGTTCATGAGACGTGGGCAATCGCGTGTCCCG 881
44   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
45 DB 390 PGGAGQKGDQGVKSSSEQGVKRGKRGGENSVVRIVGSSNRGAEVYVSGTWGTTICDD 449
46
47 QY 882 TGGATTCGTCGCCCGAGCTGCTTGGGGCGGAAGCTAACAG 922
48   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
49 DB 450 EMQNSDAIVPCRMGLGYSKRALYVAGATGQITWIDNVQCRG 490
50
51 RESULT 10
52 US-10-184-634-332
53 Sequence 332, Application US/10184634
54 Publication No. US20030068684A1
55 GENERAL INFORMATION:
56 APPLICANT: Baker, Kevin P.
57 APPLICANT: Chen, Jian
58 APPLICANT: Desnoyers, Luc
59 APPLICANT: Goddard, Audrey
60 APPLICANT: Godowski, Paul J.
61 APPLICANT: Gurney, Austin J.
62 APPLICANT: Pan, James
63 APPLICANT: Smith, Victoria
64 APPLICANT: Watanabe, Colin K.
65 APPLICANT: Wood, William I.
66 APPLICANT: Zhang, Zemin
67 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
68 TITLE OF INVENTION: ACIDS ENCODING THE SAME
69 FILE REFERENCE: P3430R1C217
70 CURRENT APPLICATION NUMBER: US/10/184,634
71 CURRENT FILING DATE: 2002-06-28

```

; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 332  
; LENGTH: 520  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-184-634-332

Query Match 3.6%; Score 40.8; DB 14; Length 520;  
Best Local Similarity 18.5%; Pred. No. 0.0098;  
Matches 74; Conservative 81; Mismatches 246; Indels 0; Gaps 0;

QY 522 ACTCATGTTCAAGTGTGACGTGAGTGAACAATCCTCGAATCCTTCTCGATGCGATGTT 581  
DB 90 DAPSFSLQSAHGEHILAQASRLQYLQQLTWVRVSHHLLQRYDNFTQNGMRKIBE 149  
QY 582 CGATCGTGTGACTTCCAAACTGCTGATCTCTGGATTTGCTGGCCACTACGTAATTCGGG 641  
DB 150 QGAPGLQGHKAMGMPGAPPGPPPAKKGAKGMDGATGSPGPGSPGVKGEAGLQGP 209  
QY 642 TGATCTGATGATGACGAAGAATCGGCGGTGAGTGTCAAGTTCCTGCTGCGATCTGAC 701  
DB 210 QGAPGKQATGTTPPGGEGSKDGLIGPKETGTGKRGDGLPSSKDRGMKDDAV 269  
QY 702 GAAATTTGATCTGGCGTGTATGAGAGTTGATGAGCGACGAAAAAGTGTGACA 761  
DB 270 MGPAGQSKGDPGRGPPGLAGFPAGKDDQGPGLGVPBPAGAVGHGAKGEPGAS 329  
QY 762 TGGCAACCTGACCTCGGAAATGCTTGTATGAGTGGCGGGAATCCACACCTT 821  
DB 330 PPARGLPGSPGATGKSKDGTGLQOQGRKSSVPBPAGVGEQSSPGLAPKGA 389  
QY 822 GGAACGTGCGGAGATTTGAGTTTGTTCGATGAGCGCGCAATCGCTGTCGCC 881  
DB 390 PQAGGKGPQGVKSGSGEGGVKGEKBERGENSVIVSSNRGRVYTSGTWGTICDD 449  
QY 882 TGGATTGCGTCCGAGCTGCTTGTGGGCAAGACTACAG 922  
DB 450 EMONSDAIVFCRMLGYSKGRALYKVYAGTGQILMDVQCRG 490

## RESULT 11

US-10-140-472-10  
; Sequence 10, Application US/10140472  
; Publication No. US20030138888A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C168  
; CURRENT APPLICATION NUMBER: US/10/140,472  
; CURRENT FILING DATE: 2002-05-06  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 10  
; LENGTH: 594  
; TYPE: PRT

; ORGANISM: Homo Sapien  
US-10-140-472-10

Query Match 3.3%; Score 37.6; DB 12; Length 594;  
Best Local Similarity 9.4%; Pred. No. 0.12;  
Matches 36; Conservative 143; Mismatches 204; Indels 0; Gaps 0;

QY 692 CGTATCTGACGAAATTCATCTGCGTTGATGAGTGTGAGTGGCAGCAAGAAA 751  
DB 195 SDEBEVLPDIVEVDVBLNDEYADLNKQATTTGMDADGVRMLRKDKERAEATKAKA 254  
QY 752 GTGTGACATGACGACCGTGCACCTGGAAATTCGTTGATGACAGTGGCGGGATC 811  
DB 255 LEEEXAMYSGRHSRQRREFREKRLRGKISPPSYARADSPYDPYKSPSSSSRSR 314  
QY 812 CACAAGCCTTGAATCTGCGGAGAAATTTGATTTGATTCATGAGAGTGGCGCAATCG 871  
DB 315 SRSPTPGEEKITFTFTSGSDDEAAAAAAGVTTGPPAPDPGPAPGNNASAR 374  
QY 872 CGTGTCCTCGGATGTCGTCGCGAGCTGCTTGGGCGCAACTACAGATGCTCAGA 931  
DB 375 RRSSSSSSSASRTSSSSSSSSRRRGGYRSGRHASRSRSRSRRRYSR 434  
QY 932 TCGTTGACAGAGAAATCTGCCCGGTCAATGTTGCATCTTCGCGCGCATGTGA 991  
DB 435 SRSRGRHRSRSGSRDGHYRSRPARCGYGPRRSRSSRSHSGDRYRGGRLRHSSRS 494  
QY 992 TGATGATTCGAGCTTGTGCGGCTTAACGCTTCGCGGAGATCCTCGACTCGTCAAGCTG 1051  
DB 495 RSMWLSFSRSLTRSHSPSPGSRSSRSRSPSPAPARKLTPAPAPVAGEL 554  
QY 1052 GCGCAGCTTGGAAATCTTCGTGA 1076  
DB 555 KTEBPAGKETGAAKVTQADASGEA 579

## RESULT 12

US-10-141-761-10  
; Sequence 10, Application US/10141761  
; Publication No. US20030148432A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C198  
; CURRENT APPLICATION NUMBER: US/10/141,761  
; CURRENT FILING DATE: 2002-05-08  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 550  
; SEQ ID NO 10  
; LENGTH: 594  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-141-761-10

Query Match 3.3%; Score 37.6; DB 12; Length 594;  
Best Local Similarity 9.4%; Pred. No. 0.12;

```

OY 752 GTGGTCACATGGCAACCGTGACACCTCGGAAATCGCTGTATGCAGATGGCGGGATC 811
Db 255 LEEBKAMYSGRRSRQRRRFRERKRLRKRKLSPEYARDSTYPIYKRSBESSSESRSR 314
OY 812 CACAGCGTTGGAACCTGTGGCGAGAAATTGATGTTGGTTGCATGGAGACGTGCGCAATCG 871
Db 315 SRSPTPRREKRTFTTSPFGSGDEFAAAAAAASGVTYTKPAPAPQPGGAPGRNMSAR 374
OY 872 CGTGTCGCCGTGCATTTCCGTGCCGACCTGTCTTGGGGCAGAACTACAGATCTCAGA 931
Db 375 RRSSSSSSSASASTSSSSSSSSSSSSRRGGGYRRGRHARSRSRMSRSRSRRSRSR 434
OY 932 TCGTTGACAGAGAAGATCTGCCCCCGGCATCATGTTGCATATCCGCGCGCATGTCGA 991
Db 435 SRSGRRHSGSGSDGHRYSRSPARRGYPFRRSRSRSGDYRRRGRLRHHSSRSR 494
OY 992 TGATGATTCGACACTGTGCGGCTTACGCTTCCTGCGCAGATCCTCGGACTCGTCGAAGCTG 1051
Db 495 RSSWSLSPRSRSLTRSRSHSPSPSGRSRSRSRSGSPSPAPRERKULTRAPASRAVERKL 554
OY 1052 GCGGCACTTGGGAAATCTTCGTGA 1076
Db 555 KTEPAAGKETGAAKVTQADASGBA 579

RESULT 14
US-10-158-790-10
; Sequence 10, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tunas, Daniel
; APPLICANT: Macanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zhenli
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C448
; CURRENT APPLICATION NUMBER: US/10/158, 790
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-158-790-10

Query Match 3.3%; Score 37.6; DB 12; Length 594;
Match Local Similarity 9.4%; Pred.No.0.12;
Matches 36; Conservative 145; Mismatches 204; Indels 0; Gaps 0;

OY 692 CGTATCGACGAATAATGCATCTGGCGTGTATGAGAGTTGTAAGTGCAGACGAATAA 751
Db 195 SDBEVIPLDIVEVDVDELNOEQVADLNKQATTYGADGDVRYRLRKDKREALALIKAKA 254
OY 752 GTGTCGACATGGCAACCGTGACACCTCGGAAATCGCTGTATGCAGATGGCGGGATC 811
Db 255 LEEBKAMYSGRRSRQRRRFRERKRLRKRKLSPEYARDSTYPIYKRSBESSSESRSR 314

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 8, 2003, 18:48:11 ; Search time 1964 Seconds  
(without alignments)  
14107.486 Million cell updates/sec

Title: US-10-007-527A-1

Perfect score: 1140

Sequence: 1 atgaccacgctgaatgcctga.....taatacgggttcgcaatga 1140

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estm:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hcc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_pro:\*  
26: em\_gss\_pig:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	3.8	767	10	BF862931 963039F08
2	43	3.8	1071	29	CC304102 CH261-221
3	41.8	3.7	609	14	CD432432 ETH1 30 H
4	41.8	3.7	885	13	BK425603 BK425603

C 5	40.8	3.6	1201	13	BX394611	BX394611
C 6	40.6	3.6	556	12	B1440943	B1440943
C 7	40.6	3.6	737	12	BW931311	BW931311
C 8	40.6	3.6	744	13	BQ442682	BQ442682
C 9	40.6	3.6	937	13	BUS91391	BUS91391
C 10	40.6	3.6	1122	11	AK084367	AK084367
C 11	40.2	3.5	1201	13	BX376097	BX376097
C 12	39.4	3.5	1201	13	BX381961	BX381961
C 13	38.8	3.4	759	9	AI663718	AI663718
C 14	38.8	3.4	365	10	CA361593	CA361593
C 15	38.2	3.4	491	12	BG406185	BG406185
C 16	38.2	3.4	515	10	B1469223	B1469223
C 17	38.2	3.4	515	10	BQ404909	BQ404909
C 18	38.2	3.4	568	9	AM185913	AM185913
C 19	38	3.3	773	12	B1555257	B1555257
C 20	38	3.3	787	29	CNS016H3	CNS016H3
C 21	38	3.3	812	9	AL695087	AL695087
C 22	37.8	3.3	686	28	BH576806	BH576806
C 23	37.6	3.3	825	29	CNS02F5W	CNS02F5W
C 24	37.6	3.3	1402	10	BG395673	BG395673
C 25	37.4	3.3	639	10	BB658123	BB658123
C 26	37.2	3.3	1201	13	BX381961	BX381961
C 27	36.8	3.2	636	12	BG820861	BG820861
C 28	36.6	3.2	242	9	AI437648	AI437648
C 29	36.6	3.2	485	10	BG650579	BG650579
C 30	36.6	3.2	595	14	CA785996	CA785996
C 31	36.6	3.2	639	10	BG415045	BG415045
C 32	36.6	3.2	1101	29	CNS016TW	CNS016TW
C 33	36.4	3.2	692	12	BMS96227	BMS96227
C 34	36.2	3.2	374	12	B1780520	B1780520
C 35	36.2	3.2	528	14	CA029424	CA029424
C 36	36.2	3.2	628	13	BQ739865	BQ739865
C 37	36	3.2	334	9	AW508862	AW508862
C 38	36	3.2	925	29	CNS0091P	CNS0091P
C 39	36	3.2	1201	13	BX444391	BX444391
C 40	35.8	3.1	292	10	BA431085	BA431085
C 41	35.8	3.1	1645	29	CNS01213	CNS01213
C 42	35.8	3.1	645	9	AL577643	AL577643
C 43	35.6	3.1	637	14	CB243032	CB243032
C 44	35.6	3.1	637	10	BB581130	BB581130
C 45	35.6	3.1	849	13	BX462111	BX462111

#### ALIGNMENTS

RESULT 1  
LOCUS BF862931  
DEFINITION 963039F08.y1 C. reinhardtii CC-1690, Stress condition I, normalized  
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF862931  
VERSION BF862931.1  
KEYWORDS GI:1225075

SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE 1 (bases 1 to 767)  
Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,  
Lefebvre, P., McDermott, J. P., Shriver, J., Slightow, C., and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 3  
Unpublished

JOURNAL CONTACT: Charles Hauser  
COMMENT DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauuser@duke.edu.  
Location/Qualifiers

#### FEATURES



```

source
1. .767
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, stress condition I,
normalized, lambda Zap II"
/notes="vector: pluescript II SK-, site 1: EcoRI, site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McMermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pluescript II SK- plasmids were excised from the lambda
Zap clones by superinfection with Exsist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

```

	Query Match	3.8%: Score 43; DB 10; Length 767;
	Best Local Similarity	48.6%: Pred. No. 0.26; Mismatches 125; Indels 0; Gaps 0;
	Matches 118; Conservative 0;	
QY	390 GGCAGCCTGGAAAAGCTGCGAACCAACGGTCGTCGTTGGCGTACCGGACGTAAGATGTACGG	449
Db	190 GGCAGCCTGCGAGAGCGGAGGGCTGGCCGCGATTATGTCGGCGCTGGAGAGCGCCAGAAACCG	249
QY	450 CTGGAGAGGATACCTGGCGCGCTGTGTGAATATCTACCGGAAAAACGGCTGGACGTCGA	509
Db	250 CGGGAGAGACTACTCGCTGGCGAGATATCGCGCGCTGAGAAAGCGTCCAGAGGGCCA	309
QY	510 CGTTCACGCGCTACTCATGTTCACATGGGACGTAAGTGAAGAACATCCTCGAATCCTTC	569
Db	310 CGTAAAGGTGAAGCCCAAGACGGGTGGGGCGCGCGGACCCCAATCTTCCGGGCGCGTGG	369
QY	570 GGATGCAATGTTGATGGGTGGACTTCCAAATCGTATCTCTGGAGATTGCTGGCGCACT	629
Db	370 GGCAGCGGTGGCGCGCGCGCGACGAGGCTCGCACTTGAGTGGATGAGGAGGCTTGGCGCC	429
QY	630 ACG 632	
Db	430 GCG 432	
RESULT 2	CC304102	1071 bp DNA linear GSS 13-MAY-2003
LOCUS	CH261-22111.Sp6.1.CH261	Gallus gallus genomic clone CH261-22111,
DEFINITION		genomic survey sequence.
ACCESSION	CC304102	
VERSION	CC304102.1	GI:30675543
KEYWORDS	GSS.	
SOURCE	Gallus gallus (chicken)	
ORGANISM	Gallus gallus	
	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Archosauria; Aves; Neognathae; Galliformes; Phasianidae;	
	Phasianinae; Gallus.	
	1 (bases 1 to 1071)	
REFERENCE	Kremetzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,	
AUTHORS	Warren,W., Graves,T., Mardis,E. and Wilson,R.	
	Gallus gallus BAC End Reads	
	Unpublished	
TTITLE	Contact: Richard K. Wilson	
JOURNAL	Genome Sequencing Center	
COMMENT	Washington University School of Medicine	
	Email: submissions@wustl.edu	
	Insert Length: 182000 Std Error: 0.00	
	Seq primer: Sp6 ATTTAGGTGACACTATAG	
	Class: BAC ends	

FEATURES		source
High quality sequence start: 27		
High quality sequence stop: 476.		
Location/Qualifiers		
1..1071		
/organism="Gallus gallus"		
/mol_type="genomic DNA"		
/strain="Red Jungle Fowl"		
/db_xref="taxon:9031"		
/clone="CH261-22111"		
/sex="female"		
/cell_line="UCD001, inbred 256"		
/clone_1b="CH261"		
/note="Vector: PTARBAC2.1; Site.1: EcoRI; CH61 Female Chicken library - For library and clone ordering information: <a href="http://www.chori.org/bacpac">http://www.chori.org/bacpac</a> "		
BASE COUNT	254 a	312 c 213 g 290 t 2 others
ORIGIN		
Query Match	3.8%;	Score 43; DB 29; Length 1071;
Best local Similarity	53.2%;	Pred. No. 0.31;
Matches 91; Conservative	0;	Mismatches 80; Indels 0; Gaps 0;
Qy	840	TGAGTTGGTTCATGGAGCGTGGGCAATCGCGTGTCCCGTGAATTGGTGGCCGAGC 899
Db	69	TGTGTGTGCTGCGTGGGAGCTGTGGACAGAGGTGAATTCACCCGATGTGTGCAAGTGTG 128
Qy	900	TGATCTTGGGGCAGAACTTAAACAATGCTCAATGTGTTGAGCAGAGAGAATCTGGCCCGGT 959
Db	129	TGGCACTGCATTCACATGCTGCTGATGTGCGTGTGGCTGGGCAAGAAAGCTTGAACATGCA 188
Qy	960	CATGGTTGCATCATTCCTGGGGGCGAATGCTGGATGATGATTCGGACTTGTGC 1010
Db	189	CCCGGTTCTGTACAGCTGTGTGTGACTCTCTGGACCGGTATGACTGAGACTGTGTC 239

RESULT 3	CD432432	609 bp	mRNA	linear	EST 03-JUN-2003
LOCUS	CD432432				
DEFINITION	ETH1_30_H03.b1.A002 Ethylane-treated seedlings Sorghum bicolor cDNA				
ACCESSION	CD432432				
VERSION	CD432432.1				
KEYWORDS	EST.				
SOURCE	Sorghum bicolor (sorghum)				
ORGANISM	Sorghum bicolor				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.				
AUTHORS	1 (bases 1 to 609) Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua Tat, N., Gonzalez, M., Lane, S., Miller, V., Nanda, F., Olaseinde, O., Baetman, A. and Pratt, L.H.				
TITLE	An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid (ACC)-treated seedlings				
JOURNAL	Unpublished				
COMMENT	Other ESTs: ETH1_30_H03.g1.A002 Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University, sequencing done in the laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA. Seq primer: Sng3 (CGACCTGCAGCTCGACGACA)				

[illegible]

JOURNAL  
COMMENT

Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: sege@genoscope.cns.fr. Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9540.f. For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODC018BD100P1cluster=9540.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paraday Avenue Genoscope sequence ID : CSODC018BD100P1.

## FEATURES

## source

```
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODC018Y20"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (drr)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
```

## BASE COUNT

192 a 357 c 358 g 218 t

## ORIGIN

## Query Match

Best Local Similarity 37.6%; Score 40.8; DB 13; Length 1201;  
Matches 70; Conservative 34; Mismatches 82; Indels 0; Gaps 0;

```
QY 782 AATGCTGTGATGATGAGCGGGGATCCACAAAGCGTTGAGTGGCGGAAATTG 841
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1153 AAGAVSCTGATGCTGTGGAAKXGKGCCCACCCKTGKMGCCMMGGCAATKDS 1094
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 842 AGTTTGTTTCATGGAGCTGCGGCAATGCGCTGCTCCCGTGAATTGGTCCGAGCTG 901
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1093 CCCCTGCGYGTGGCCCTKCCYSCCCAGACAKGATGKKGATTKSSCACCGAGCTK 1034
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 902 GTCTGGGGGAGAACTAAGATGCTGAGATGTTGAGAGAGAAATCTGCCCCGCTCA 961
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 1033 GSCYGAAGKGTASTGACATGTGACYSGCCAASAWATGAGCCGCTGGGGAGCTGSCC 974
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 962 TGGTGG 967
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 973 TGGGAG 968
```

## RESULT 6

## BI440943/c

LOCUS 556 bp mRNA linear EST 12-MAR-2002

## DEFINITION

musculus CDNA clone IMAGE:5659863 5' similar to SN:PTX3\_MOUSE  
P48759 PENTAXIN-RELATED PROTEIN PTX3 PRECURSOR ; mRNA sequence.

## ACCESSION

## BI440943

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

1 (bases 1 to 556)  
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, J., Clifton, S.,  
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blaisdell, A.,  
Schmitt, A., Theising, B., Ritter, E., Komko, I., Bennett, J., Cardenas,  
, W., Gibbons, M., McCann, R., Cole, R., Tsagarelis, R., Williams, T.,  
Jackson, Y., and Bowers, Y.

## TITLE

## JOURNAL

## COMMENT

Unpublished  
Other ESTs: ic78e08.xl  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138

Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmellon@iobp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Juliana Brown  
(brown@fas.harvard.edu)

Possible reversed clone: similarity on wrong strand  
MGI:1946189 This sequence now available from the IMAGE consortium,  
for clone orders contact: info@image.llnl.gov

High quality sequence stop: 437.

## FEATURES

## source

```
1..556
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5659863"
/sex="both for embryonic & newborn, male for adult and
adult islet"
/dev stage="Embryonic day 10.5, B12.5, B16.5, newborn,
adult, mixed"
/lab host="MDH10B"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
NI-M5T"
```

/note="Vector: pSPORT1, Site 1: Not I, Site 2: Sal I, Five  
libraries representing B10.5/12.5 pancreatic bud, B16.5  
pancreas, newborn pancreas, adult pancreas, and adult  
islets of Langerhans were separately constructed using  
SuperScript Plasmid Library kit (Life Technologies). cDNA  
was made by oligo-dT priming and size-selected by column  
fractionation. Libraries were amplified once on solid  
support and plasmid DNA from each library was prepared  
and mixed in equal amounts. The mixed library DNA was  
normalized by method #4 from Bonaldo, Lemmon, and Soares  
1996 Genome Research 6:791-806; 0.5 microgram  
single-stranded mixed library plasmid DNA was mixed with  
5 micrograms PCR product representing mixed library  
inserts and hybridized to an EcoT of 6. Single-stranded  
(unhybridized) plasmids were isolated by hydroxyapatite  
chromatography and used to make this library."

## BASE COUNT

82 a 191 c 171 g 112 t

## ORIGIN

## Query Match

Best Local Similarity 48.5%; Score 40.6; DB 12; Length 556;  
Matches 112; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

```
QY 696 TCTGACGAATTTGATCTGCGCTGTGTATGAGAGTTGTGTGCGGCAAAAATGCG 755
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 537 TCGTGAAGACTCTGAGAGCTGTGCGGAGAGCTGGCGGCGGCGGCGCATGTGGGA 478
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 756 TCAGATGCGACCGGTGACCCCTGGGAATAGCTGTGTGATGAGAGCGGCGGATCCACA 815
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 417 TGCAGAGAGCGCGTGAAGCCGAGCTCAGAGCTGAGCGCGCTGGAAGACGCGAGCGCGC 358
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 477 GCGCGTGGCAGCGCGGTGCGCCCGCAGACGCAAGCTGTGTGAGAGCGCGCTGC 418
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 816 AGCGTTGGAAGTGTGCGGAGAAATTTGATTGTGTGATGAGAGCGTGGGCAATGCGCTG 875
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY 876 GTCCCGTGAATTTGCGTCCGAGCTGTGTGTTGGGCGGAGAACTAACAGATGC 926
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 357 GACCGAGGCGCAGTGTGCTGAGCGCTGTGTGAGAACTGCGGC 307
   |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
```

## RESULT 7

## BM933131/c

LOCUS 737 bp mRNA linear EST 13-MAR-2002  
DEFINITION UT-M-BH3-btw-a-11-0-UT.r1 NIH\_BMAP\_M\_S4 Mus musculus cDNA clone

UI-M-BH3-brw-a-11-0-UI 5', mRNA sequence.

ACCESSION  
BM333131 GI:19392283

VERSION  
EST

KEYWORDS  
Mus musculus (house mouse)

SOURCE

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 737)

Bonaldi,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene  
discovery

Genome Res. 6 (9), 791-806 (1996)

JOURNAL  
97044477

MEDLINE  
PUBMED

COMMENT

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: m85t@mail.nih.gov

CDNA Library preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 REVERSE.

#### FEATURES

Location/Qualifiers

1..737

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UI-M-BH3-brw-a-11-0-UI"

/dev\_stage="27-32 days"

/lab\_host="DH10B (Life Technologies)"

/clone\_id="NIH\_BMAP\_M\_S4"

/note="Vector: pT713D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The

NIH\_BMAP\_M\_S4 library is a subtracted library of a series,

ultimately derived from a mixture of individually tagged

normalized libraries from ten regions of the mouse brain

(cerebellum, brain stem, olfactory bulbs, hypothalamus,

cortex, amygdala, basal ganglia, pineal gland, striatum,

hippocampus) after a series of subtractions to reduce the

representation of cDNAs from which ESTs had already been

generated. The following serially subtracted libraries

were generated in this process: NIH\_BMAP\_M\_S4,

NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,

NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library

(NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified

CDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and

NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived

was used as a driver in a hybridization with a pool of

the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1

libraries in the form of single-stranded circles. The

remaining single-stranded circles (subtracted library)

was purified by hydroxyapatite column chromatography

converted to double-stranded circles and electroporated

into DH10B bacteria (Life Technologies) to generate the

NIH\_BMAP\_M\_S4 library. This procedure has been previously

described (Bonaldi, Lennon and Soares, Genome Research

6:791-806, 1996)

BASE COUNT  
ORIGIN  
122 a 246 c 218 g 149 t 2 others

Query Match 3.6%; Score 40.6; DB 12; Length 737;  
Best Local Similarity 48.5%; Pred. No. 1.3;  
Matches 112; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 696 TCTGACGAATTCGATCTGCGCTGTGATGAGCTGTGATGCGGACGAGAAAGTGG 755

Db 564 TCCGTGAGAGCTGACGCGCTGCGGCGAGACTGGGCGGCTGCGCGCGCATGCGCA 505  
QY 756 TCACATGCGCAACCGTCGACCTGCGGAATGCTGTGTGATGACGATGGGCGGGATCCACA 815  
Db 504 GCGCGTGCAGACCGCGTGGCGCCCGCAGACCGCAGGCTGTGCGGCGCGCTGC 445  
QY 816 AGCCTTGGAATCTGTGCGGAGATTGAGTTTGTTCATGAGGACGTCGCGCATCGCGTG 875  
Db 444 TGCAGAGAGACCGTGAAGCGAGCTTACGCGCTGGCGCGCTGAGAGAGCGGAGCGCGC 385  
QY 876 GTCCCGTGAATGCTGCGCGGAGCTGCTTGGGCGGAACTAACAGATGC 926  
Db 384 GACCCGAGGCGCAGCTGCTGCTAGCGCTGTGCGAGAACTCGCGC 334

#### RESULT 8

LOCUS

BM442682 744 bp mRNA linear EST 29-MAY-2002

DEFINITION UI-M-EV0-bxf-a-13-0-UI.r1 NIH\_BMAP\_EV0 Mus musculus CDNA clone

IMAGE:5707284 5', mRNA sequence.

ACCESSION BM442682

VERSION BM442682.1 GI:21245794

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.

1 (bases 1 to 744)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

CONTACT: Robert Strauberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa

CDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..744

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:5707284"

/tissue\_type="whole brain"

/dev\_stage="embryo 15.5 dpc"

/lab\_host="DH10B (11 phage resistant)"

/clone\_id="NIH\_BMAP\_EV0"

/note="Organ: Brain; Vector: pYX-Aac; Site 1: EcoR I;

Site 2: Not I; The library was constructed according to

Bonaldi, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured mRNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was size selected according to mRNA size fraction,

ligated with EcoR I adaptor, digested with Not I, and then

cloned directionally into pYX-Aac vector. The library tag

sequence located between the Not I site and the polyA tail

, is GTGCGTGA. This library was created for the

University of Iowa Mouse Brain Molecular Anatomy Project

(BMAP): 'Gene Discovery in the Developing Mouse Nervous

System', supported by National Institutes of Mental Health

(NIMH), Hemm Chin, Ph.D., program coordinator."

BASE COUNT

140 a 212 c 244 g 146 t 2 others

Query Match

3.6%; Score 40.6; DB 13; Length 744;



Yunshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
JOURNAL  
NATURE 409 (6821), 685-690 (2001)  
MEDLINE  
21085660  
PUBMED  
11217851

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
NATURE 420, 563-573 (2002)  
6 (bases 1 to 1122)

ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., BONO, H., CARINCI, P., FUKUDA, S., FURUKO, M., HANAGAKI, T., HARA, A., HASHIZUME, W., HAYASHIDA, K., HAYATSU, N., HIRAMOTO, K., HIRAKO, T., HIROZANE, T., HORI, F., IMOTANI, K., ISHII, Y., ITOH, M., KAGAWA, I., KASUKAWA, T., KATO, H., KAWAI, D., KOJIMA, Y., KONDO, S., KONNO, H., KODA, M., KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., MURATA, M., NAKAMURA, M., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OHSAKI, N., OKAZAKI, Y., SAITO, K., SAITOH, H., SAKAI, C., SAKAI, K., SAKAZUME, N., SANO, H., SASAKI, D., SHIBATA, K., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKIHARA, S., TAKEDA, Y., TANAKA, T., TOMARU, A., TOYA, T., YASUNISHI, A., MURAMATSU, M., and HAYASHIZAKI, Y.  
Direct Submission  
Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: http://genome.resgsc.riken.go.jp, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.resgsc.riken.go.jp/  
URL: http://fantom.resgsc.riken.go.jp/.  
Location/Qualifiers  
1..1122  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM:DB:D230032H18"  
/db\_xref="taxon:10090"  
/clone="D230032H18"  
/tissue\_type="eyeball"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
1..1122  
/note="Pentaxin related gene (MGI:104641, GB|NM\_008987, evidence: BLASTN, 100%, match=404)"

BASE COUNT  
244 a 327 c 292 g 259 t

ORIGIN  
misc\_feature  
Query Match 3.6%; Score 40.6; DB 11; Length 1122;  
Best Local Similarity 48.5%; Pred. No. 1.6;  
Matches 112; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Db 696 TCTGACGAAATTGTCATGCGCTGGTGTGAGAGTTGAGTGGCGAAGGAAAGTGG 755  
Db 647 TCCGTGAGAGCTCAGCGCTGCGGCGAGAGCTGCGGCGCGGCGGCGATGGGGA 588  
Qy 756 TCGACATGGAACGCTGACCGCTGGGAATGCTGTTGATGACATGGCGGGATCCACA 815  
Db 587 GCGCGTGGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 528  
Qy 816 AGCGTTGAACTGTGCGGAGAAATTGATTGTTGATGATGGGAGCGTGGGCAATCGGCG 875  
Db 527 TCGAGAGAGCGCGTGAAGCGAGCGCTCAGCGCTGCGCGCGCGCGCGCGAGGCGCGCGC 468

Qy 876 GTCCGTGATGTCGTCGCCGAGCTGATGTTGGGCGAGAACTACAGATGTC 926  
Db 467 GACCCGAGCGCAGAGCTGCTGAGCGCTGATGCTGCTGAGAGAACTGCGGC 417

RESULT 11  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segret@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2866.f  
Contact: Feng Liang Email: fliang@lifestech.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue genoscope sequence ID: CS0DC022B606P1.  
Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC022YMI2"  
/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT  
79 a 152 c 66 g 270 t

ORIGIN  
Query Match 3.5%; Score 40.2; DB 13; Length 1201;  
Best Local Similarity 8.8%; Pred. No. 2.2;  
Matches 54; Conservative 242; Mismatches 317; Indels 2; Gaps 1;

Qy 165 TGTGACCATTTGTCAACGCTCGAAGGTTCTGATTCGAGAGCGCTGCTCGGAAA 224  
Db 431 TMTTNNMCMCTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 490  
Qy 225 GGGCTGATCTGCGCCCTGCTGCGGAGAAAGTGGTGCATCGTGCAGGAATTTC 284  
Db 491 NKKNNKKNN 550  
Qy 285 TCAAGTTGTCATCATCACTCGGAGCTGATCTGTTGCGATGGTGAAGATGACCATGCG 344  
Db 551 KKKKKNN 610  
Qy 345 CCATACAGCTGTCACGCGCTCCAGACCTATGAGCTGATTCGCGAGCGCTGGAAGC 404  
Db 611 CKKKNN 670  
Qy 405 TCGACCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 462  
Db 671 MKNN 730  
Qy 463 GTGCGCGCTTTGAATCACTGACGAGAAAGGCTGGGACGTCACGCTTCAAGCGCTA 522  
Db 731 KKKKKNN 790



Dib 1081

ORIGIN



Query Match 3.4%; Score 38.8; DB 9; Length 759;  
 Best Local Similarity 52.1%; Pred. No. 4.4;  
 Matches 85; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 764 GCAACCGTGCACCTCGGAAATGCTGTGATGAGTGGCGGAGATCCAAAGCTTGG 823  
 DB 602 GCAGCGGTGGCCCNCGAGACGAGCTGTGTGGGGCGCTGAGCGCTGCTGCAAGAG 543  
 QY 824 AACGTGGCGAGAAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 883  
 DB 542 AGCCCTTACGCGAGCTTACGCTGCGCGCTGAGAGAGCGGAGCGCGGCAACCGAG 483  
 QY 884 GATTGCGTGGCCGAGCTGTGCTTGGGCGAAGTACAGATGC 926  
 DB 482 GCGACAGTGCCTGGCTGAGCGCTGTGTGAGAGAACTGCGGC 440

RESULT 14  
 CA361593 777 bp mRNA linear EST 06-NOV-2002  
 LOCUS 63452 NCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT6015\_A\_H08 5',  
 DEFINITION mRNA sequence.

ACCESSION CA361593.1 GI:24667466  
 VERSION CA361593.1  
 KEYWORDS EST.  
 ORGANISM Oncorhynchus mykiss (rainbow trout)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 Protopranchiopherygii; Salmoniformes; Salmonidae; Oncorhynchus.

REFERENCE 1 (bases 1 to 777)  
 AUTHORs Rexroad, C.E. and Keefe, J.W.  
 TITLE Sequence analysis of a rainbow trout normalized cDNA library  
 JOURNAL Unpublished  
 COMMENT Contact: Rexroad CE  
 USA, ARS, National Center for Cool and Cold Water Aquaculture  
 11876 Leestown Road, Kearneysville, WV 25430, USA  
 Tel: 304 724 8340 X2129  
 Fax: 304 725 0351  
 Email: crexroad@nccwa.ars.usda.gov

FEATURES  
 source  
 1..777  
 Location/Qualifiers  
 /organism="Oncorhynchus mykiss"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8022"  
 /clone="1RT6015\_A\_H08"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_1lb="NCCWA 1RT"  
 /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
 library made from pooled tissue from brain, gill, liver,  
 spleen, muscle, and kidney."

BASE COUNT 199 a 221 c 188 g 169 t  
 ORIGIN

Query Match 3.4%; Score 38.8; DB 14; Length 777;  
 Best Local Similarity 54.1%; Pred. No. 4.4;  
 Matches 79; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 282 TTCTCAATGTTGCTCATCACTGGAGCTGATCTGTGCAATGTGACGATGACAT 341  
 DB 285 TTCTCAATGTTGCTCATCACTGGAGCTGATCTGTGCAATGTGATGATGAGAAC 344  
 QY 342 GCGCCATACAGCTGTGTCAGCGCTCAAGCCTATGAGACTTGCTGGCAGCTGAA 401  
 DB 345 TCCCTTTCGGCTCCGCGAGGCTCCCAAGGCTTAGGGGTCTATGACAGAGATTTCT 404  
 QY 402 AGCTGCAACCAAGCTGCTGTGGC 427

DB 405 CGCTGCTCCCGCCAGCAAGTCCGTGAC 430

RESULT 15  
 BG406185/c 365 bp mRNA linear EST 28-NOV-2001  
 LOCUS sac36f04.y1 Gm-cl051 glycine max cDNA clone GENOME SYSTEMS CLONE  
 DEFINITION ID: Gm-cl051-4472 5' similar to TR:004253 004253 T10M13.14.; mRNA  
 sequence.

ACCESSION BG406185  
 VERSION BG406185.1 GI:33312534  
 KEYWORDS EST.  
 ORGANISM Glycine max (soybean)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseolaeae;  
 Glycine.

REFERENCE 1 (bases 1 to 365)  
 AUTHORs Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna,  
 A., Bolla, B., Maira, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
 Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers,  
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk,  
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann,  
 R., Waterston, R. and Wilson, R.

COMMENT Contact: Shoemaker R/Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu

FEATURES  
 source  
 1..365  
 Location/Qualifiers  
 /organism="Glycine max"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl051-4472"  
 /tissue\_type="floral meristematic mRNA"  
 /lab\_host="DH10B"  
 /clone\_1lb="Gm-cl051"  
 /note="Vector: pBluescript II SK+, Site 1: EcoRI; Site 2:  
 XhoI; The cDNA library was constructed from floral  
 meristematic mRNA provided by Dr. Halina Knap of Clemson  
 University. Complementary DNA was synthesized from mRNA  
 using a primer consisting of a poly(dT) sequence with a  
 XhoI restriction site. EcoRI adapters were ligated to the  
 blunt-ended cDNA fragments followed by XhoI digestion. The  
 cDNA fragments were directionally cloned into the  
 EcoRI-XhoI restriction site of the pBluescript vector. The  
 ligated cDNA fragments were transformed into DH10B host  
 cells (GibcoBRL). This library was constructed in the  
 laboratory of Dr. Randy Shoemaker."

BASE COUNT 75 a 154 c 57 g 78 t  
 ORIGIN

Query Match 3.4%; Score 38.2; DB 10; Length 365;  
 Best Local Similarity 53.7%; Pred. No. 4.4;  
 Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 681 AGTTTCGCTGCGATCTGACGAAATTCATCTGCGCTGTGATGAGGTTGATGAG 740  
 DB 261 AGACCTTGAAGTTCAATGAGGAAAGTGATTTGACAGAGGATTCAGAGTTGCTGG 202  
 QY 741 CGACGAAATAGTGTGACATGGAACCGGACCCCTGGGAAATGCTGTGATGACGT 800  
 DB 201 AATCAGGAAGTGAAGATGTTCTCCGGGCACTGGGAGTGTGGAGATGAGTGA 142

Oy 801 GGGGGGATCCACAAGCTTGGACT 827  
|||  
Db 141 CGGGCGCATGAGCCCGGAGGTAAT 115

Search completed: November 8, 2003, 21:28:16  
Job time : 1969 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 22, 2003, 09:04:13 / Search time 1957 Seconds

(without alignments)  
14157.947 Million cell updates/sec

Title: US-10-007-527A-1  
Perfect score: 1140  
Sequence: 1 atgacagcagcgaagtcgtga.....taatacgcgtccgaagtrga 1140

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estlm:\*  
3: em\_estlm:\*  
4: em\_estlm:\*  
5: em\_estlm:\*  
6: em\_estlm:\*  
7: em\_estlm:\*  
8: em\_estlm:\*  
9: gb\_est1:\*  
10: gb\_est1:\*  
11: gb\_est1:\*  
12: gb\_est1:\*  
13: gb\_est1:\*  
14: gb\_est1:\*  
15: em\_estlm:\*  
16: em\_estlm:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_hum:\*  
20: em\_gss\_hum:\*  
21: em\_gss\_hum:\*  
22: em\_gss\_hum:\*  
23: em\_gss\_hum:\*  
24: em\_gss\_hum:\*  
25: em\_gss\_hum:\*  
26: em\_gss\_hum:\*  
27: em\_gss\_hum:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	3.8	767	10	BF862931 963039P08
2	43	3.8	1071	29	CC304102 CH261-221
3	41.8	3.7	609	14	CD432432 ETH1_30.H
4	41.8	3.7	885	13	BX425603 BX425603

C 5	40.8	3.6	1201	13	BX394611 BX394611
C 6	40.6	3.6	556	12	B1440943 B1440943
C 7	40.6	3.6	737	12	BM933131 BM933131
C 8	40.6	3.6	744	13	BQ442682 BQ442682
C 9	40.6	3.6	937	13	B0591391 B0591391
C 10	40.6	3.6	1122	11	AK084367 AK084367
C 11	40.2	3.5	1201	13	BX376097 BX376097
C 12	39.4	3.5	1201	13	BX381961 BX381961
C 13	38.8	3.4	759	9	AL663718 AL663718
C 14	38.8	3.4	777	14	CA351593 CA351593
C 15	38.2	3.4	365	10	BG406185 BG406185
C 16	38.2	3.4	491	12	B1469223 B1469223
C 17	38.2	3.4	515	10	BG404909 BG404909
C 18	38.2	3.4	568	9	AM185913 AM185913
C 19	38	3.3	773	12	B1555257 B1555257
C 20	38	3.3	787	29	CNS016H3 CNS016H3
C 21	38	3.3	812	9	AL695087 AL695087
C 22	37.8	3.3	686	28	BH576806 BH576806
C 23	37.6	3.3	825	29	CNS02F5M CNS02F5M
C 24	37.6	3.3	1402	10	BG395673 BG395673
C 25	37.4	3.3	639	10	BG658123 BG658123
C 26	37.2	3.3	1201	13	BX381961 BX381961
C 27	36.8	3.2	636	12	BG820861 BG820861
C 28	36.6	3.2	242	9	AL437648 AL437648
C 29	36.6	3.2	485	10	BG650579 BG650579
C 30	36.6	3.2	595	14	CA785996 CA785996
C 31	36.6	3.2	639	10	BG415045 BG415045
C 32	36.6	3.2	1101	29	CNS016H3 CNS016H3
C 33	36.4	3.2	692	12	BM596227 BM596227
C 34	36.2	3.2	374	12	B1780520 B1780520
C 35	36.2	3.2	528	14	CA029424 CA029424
C 36	36.2	3.2	628	13	BQ739865 BQ739865
C 37	36	3.2	334	9	AM508862 AM508862
C 38	36	3.2	925	29	CNS0091P CNS0091P
C 39	36	3.2	1201	13	BX444391 BX444391
C 40	35.8	3.1	292	10	BR431085 BR431085
C 41	35.8	3.1	645	29	CNS012T3 CNS012T3
C 42	35.8	3.1	1163	9	AL577643 AL577643
C 43	35.6	3.1	317	14	CB243032 CB243032
C 44	35.6	3.1	637	10	BS581130 BS581130
C 45	35.6	3.1	849	13	BX462111 BX462111

## ALIGNMENTS

RESULT 1  
BF862931  
LOCUS 963039P08.y1 C. reinhardtii CC-1690. Stress condition I, normalized  
DEFINITION 'lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF862931.1 GI:12253075  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Chlamydomonas reinhardtii  
Chlamydomonas reinhardtii  
Eukaryota, Viridiplantae, Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadales.

REFERENCE  
AUTHORS  
TITLE  
Lefebvre, P., McDermott, J.P., Shreeve, J., Sillfow, C. and Stern, D.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants; project phase 3

JOURNAL  
COMMENT  
Unpublished  
Contact: Charles Hauser  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers

## FEATURES











Best Local Similarity 48.5%; Pred. No. 1.3; Mismatches 119; Indels 0; Gaps 0;  
Matches 112; Conservative 0;

QY 696 TCACGAGAAATTCATCTGCGCTTGTATGAGATTGGTGGAGAGAGAAAGTGC 755  
DB 304 TCCGTGAGAGCTGCAAGCGGCTGCGGAGAGCTGGGGCGGCTGGGGCGGAGTGC 363  
QY 756 TCACATGCGCAACCGTGCACCTCTGGAAATCGCTGTATGACAGTGGCGGAGTCCACA 815  
DB 364 GGCCTGCGCAGCGGTGGCGCCCGCAGACGCCAGCTGGTGGCGGGCGCTGGAGCCGCTGC 423  
QY 816 AGCGTTGAACTGTGCGCGGAATTTGATTGTGTGATGAGGAGCGTCCGAGCAATCGCGG 875  
DB 424 TGCAGAGAGCGCTGACGCGAGCTTCAAGCTGCGCGCTGGAGAGACGCGAGCGCGGC 483  
QY 876 GTCCCGTGGATTGCTGCGCCGAGCTGCTTGGGAGAGAACTAACAGATGC 926  
DB 484 GACCCGAGCGACAGTGCCTGCGCTGAGCGCGCTGTCTGAGAGAACTGCGGC 534

RESULT 9  
BU591391/c 937 bp mRNA linear EST 07-JAN-2003  
LOCUS AGENCOURT\_8783384 NIH\_MGC\_137 Mus musculus cDNA clone IMAGE:6431460  
DEFINITION 5', mRNA sequence.  
ACCESSION BU591391  
VERSION BU591391.2 GI:27538019  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 937)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
On Sep 19, 2002 this sequence version replaced gi:23243003.  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Gerard Gradwohl (PNAS 97 P1607-1611, 2000)  
cDNA Library Preparation: Catherine Lee, Endocrine Pancreas Consortium  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Plate: IRBD16 row: f column: 01  
High quality sequence stop: 660.

FEATURES  
source Location/Qualifiers  
1..937  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6431460"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_137"  
/note="Organ: pancreas; Vector: pSPORT1; site: 1; Salt; Site 2: NCI; Library consists of a pool of clones rearrayed from the following libraries: Melton normalized mixed mouse pancreas 1 NI-MMS1, Amplified Melton mouse islets 1 MSL-A, and Kaestner nrg3 wt. Clones rearrayed in the laboratory of K. Kaestner (University of Pennsylvania). Note: this is a NIH MGC Library."

BASE COUNT 183 a 293 c 249 g 212 t  
ORIGIN  
Query Match 3.6%; Score 40.6; DB 13; Length 937;  
Best Local Similarity 48.5%; Pred. No. 1.5;  
Matches 112; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 696 TCACGAGAAATTCATCTGCGCTTGTATGAGATTGGTGGAGAGAGAAAGTGC 755  
DB 304 TCCGTGAGAGCTGCAAGCGGCTGCGGAGAGCTGGGGCGGCTGGGGCGGAGTGC 363  
QY 756 TCACATGCGCAACCGTGCACCTCTGGAAATCGCTGTATGACAGTGGCGGAGTCCACA 815  
DB 364 GGCCTGCGCAGCGGTGGCGCCCGCAGACGCCAGCTGGTGGCGGGCGCTGGAGCCGCTGC 423  
QY 816 AGCGTTGAACTGTGCGCGGAATTTGATTGTGTGATGAGGAGCGTCCGAGCAATCGCGG 875  
DB 424 TGCAGAGAGCGCTGACGCGAGCTTCAAGCTGCGCGCTGGAGAGACGCGAGCGCGGC 483  
QY 876 GTCCCGTGGATTGCTGCGCCGAGCTGCTTGGGAGAGAACTAACAGATGC 926  
DB 484 GACCCGAGCGACAGTGCCTGCGCTGAGCGCGCTGTCTGAGAGAACTGCGGC 534

DB 562 TCCGTGAGAGAGCTGACGCGCTGCGGAGAGCTGCGGCGGCTGCGCGGAGTGC 503  
QY 756 TCACATGCGCAACCGTGCACCTCTGGAAATGCTGTTGATGAGTGGCGGAGATCCACA 815  
DB 502 GGCCTGCGCAGCGGTGGCGCCCGCAGACGCCAGCTGGTGGCGGGCGCTGGAGCGCTGC 443  
QY 816 AGCGTTGAACTGTGCGCGGAATTTGATTGTGTGATGAGGAGCGTCCGAGCAATCGCGG 875  
DB 442 TGCAGAGAGCGCTGACGCGAGCTTCAAGCTGCGCGCTGGAGAGACGCGAGCGCGGC 383  
QY 876 GTCCCGTGGATTGCTGCGCCGAGCTGCTTGGGAGAGAACTAACAGATGC 926  
DB 382 GACCCGAGCGACAGTGCCTGCGCTGAGCGCGCTGTCTGAGAGAACTGCGGC 332

RESULT 10  
AK084367/c 1122 bp mRNA linear HTC 05-DEC-2002  
LOCUS Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length  
DEFINITION full insert sequence.  
ACCESSION AK084367  
VERSION AK084367.1 GI:26101923  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636  
2  
Carninci, P., Shibata, Y., Hayasu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
20499374  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multitepicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Koehn, P., Lewis, S., Matsuo, T., Nikaido, I., Pesole, G., Quackenbush, J., Schiraldi, J. M., Staudt, R., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Guerinich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaeets, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Winding, L.,







Qy 801 GGGCGGGATCCACAGCGTTGAACT 827  
|||  
Db 141 CGGCGGCGATGAGAGCCGCGAGGTAAT 115  
|||

Search completed: August 22, 2003, 10:35:38  
Job time : 1962 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using fixme\_plus\_p2n model

Run on: August 22, 2003, 09:13:28 ; Search time 3108 seconds  
(without alignments)  
4988.658 Million cell updates/sec

Title: US-10-007-527a-2  
Perfect score: 2005  
Sequence: 1 MTSVSAHLSGKDRPVLVS.....HLHYRLPADVPRPPIISVRK 379

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frames+ p2n.model -DEV=x1h  
-O=/cgn2\_1/USPTO\_SPOOL/US10007527/runat 20082003 165151 15366/app query.fasta\_1.519  
-DB=GenEmbl -GEMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPTC=0 -LOOPTXT=0  
-INITs-bits -STRPT=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10007527@cgn 1.1.2312 @runat 20082003 165151 15366 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_da.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_oz.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_scs.\*  
28: em\_un.\*

29: em\_vl.\*  
30: em\_hcg\_hum.\*  
31: em\_hcg\_in.\*  
32: em\_hcg\_other.\*  
33: em\_hcg\_mus.\*  
34: em\_hcg\_pln.\*  
35: em\_hcg\_rod.\*  
36: em\_hcg\_mam.\*  
37: em\_hcg\_vrt.\*  
38: em\_sy.\*  
39: em\_hcgo\_hum.\*  
40: em\_hcgo\_mus.\*  
41: em\_hcgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2005	100.0	1140	6	AX548642	AX548642 Sequence
2	2005	100.0	6334	1	AY178757	AY178757 Rhodococ
3	2005	100.0	6334	6	AX348646	AX348646 Sequence
4	2005	100.0	9652	6	AX548648	AX548648 Sequence
5	2005	100.0	11241	6	AX548647	AX548647 Sequence
6	2005	100.0	11241	12	AY180162	AY180162 Shuttle v
7	609.5	30.4	2439	1	APU83788	UB3788 Actinomyc
8	569	28.4	3540	1	AY150274	AY150274 Propionib
9	497	24.8	2051	1	FFR6662	AY006662 Propionib
10	453	22.6	2661	1	SLMG33A	X69872 S.lavendula
11	452.5	22.6	9367	1	SNAA23257	AY172684 Coryneb
12	425.5	21.2	4603	1	AY172684	AY172684 Coryneb
13	399.5	19.9	2297	6	E17316	E17316 Bifidobac
14	392.5	19.6	5250	1	AF085719	AF085719 Bifidobac
15	374.5	18.7	5804	7	AF428260	AF428260 Propionib
16	360.5	18.0	12855	1	AF533985	AF533985 Streptomy
17	346.5	17.3	10992	1	STMPN22	D14281 Streptomyce
18	345	17.2	3748	12	AB032347	AB032347 Cloning v
19	345	17.2	5128	1	AB084167	AB084167 Pseudomon
20	345	17.2	5128	1	CCA308231	AY072037 Expressio
21	342.5	17.1	4109	1	AY072037	AY072037 Expressio
22	337	16.8	5349	12	AY072037	AY072037 Expressio
23	337	16.8	6342	12	AY072037	AY072037 Expressio
24	337	16.8	6607	12	AY072037	AY072037 Expressio
25	337	16.8	8058	12	AY072038	AY072038 Expressio
26	337	16.8	8215	12	P1414669	AJ414669 Cloning v
27	337	16.8	8830	1	PI1010G	M21778 Plasmid p10
28	333.5	16.6	2297	1	AF273219	AF273219 Pseudomon
29	326	16.3	3147	12	AF092035	AF092035 Integrati
30	326	16.3	4447	1	AF092037	AF092037 Coryneb
31	326	16.3	4447	6	ARI173114	ARI173114 Sequence
32	326	16.3	4447	6	BD106940	BD106940 Method fo
33	326	16.3	4447	6	BD133985	BD133985 Arginine
34	326	16.3	4457	1	BPLASMI	X03987 Brevibacter
35	326	16.3	5135	12	SVE133193	AY133193 Shuttle v
36	326	16.3	5216	12	SVE133194	AY133194 Shuttle v
37	326	16.3	6241	12	AF092036	AF092036 Shuttle v
38	326	16.3	6272	12	AF091268	AF091268 Shuttle v
39	326	16.3	6398	12	AF092931	AF092931 Shuttle v
40	326	16.3	6601	12	SVE133195	AY133195 Shuttle v
41	318	15.9	4447	6	ARI173115	ARI173115 Sequence
42	318	15.9	4447	6	BD106941	BD106941 Method fo
43	318	15.9	4447	6	BD133986	BD133986 Arginine
44	315	15.7	7212	1	AF327371	AF327371 Streptococ
45	296	14.8	4448	1	COGAM30	DD0038 Corynebacte

RESULT 1

ALIGNMENTS

AX548642  
 LOCUS AX548642 1140 bp DNA linear PAT 27-NOV-2002  
 DEFINITION Sequence 1 from Patent WO02055709.  
 ACCESSION AX548642  
 VERSION AX548642.1 GI:25813612  
 KEYWORDS  
 SOURCE Rhodococcus erythropolis  
 ORGANISM Rhodococcus erythropolis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Nocardiaceae; Rhodococcus.  
 REFERENCE  
 1 Brummel, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.  
 Rhodococcus cloning and expression vectors  
 Patent: WO 02055709-A1 18-Jul-2002;  
 JOURNAL E.I. DU PONT DE NEMOURS AND COMPANY (US)  
 Location/Qualifiers  
 source 1..1140  
 /organism="Rhodococcus erythropolis"  
 /mol\_type="genomic DNA"  
 /strain="AN12"  
 /db\_xref="taxon:1833"  
 BASE COUNT 241 a 287 c 347 g 265 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7.8e-119 Length: 1140  
 Score: 2005.00 Matches: 379  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-10-007-527a-2 (1-379) x AX548642 (1-1140)

QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLysAspArgProPheValLeuValSer 20  
 Db 1 ATGACCAAGCTTAAGCTTGAATTCGAGCAACGCGCTCCGCTCTGCTGCTG 60  
 QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSer 40  
 Db 61 TCCGTAAGCCGGGATCGGATCGGACGACATCGACCCAACTTCAAAATACACAGCTCA 120  
 QY 41 GluThrPheAsnAlaCysGlyArgProLysSerGlyValaGlnGlyValThrIleValAsn 60  
 Db 121 GAAACATTTAAAGCTGCTGCGCGCGGCGGATTTCTGCGGACGCGTGTGACCATTTGTAAC 180  
 QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyValThrIleCysPro 80  
 Db 181 GGTCCGAAAGGTTCTGGAATTCGAGAGGCTTCTGCTCGGAAAGGCTGGAATCTGCCCC 240  
 QY 81 CysCysAlaGlyLysValaGlyAlaHisArgAlaAspGluIleSerGlnValaAlaHis 100  
 Db 241 TCGTGTGCGGAAAGAGTGGTGCATCGTGCAGCAAAATTTCTCAAGTTGTTGCTCAT 300  
 QY 101 GlnLeuGlyThrGlySerValaAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120  
 Db 301 CAACCTCGGAGCTGATCTGTTGCGATGTGACGATGACCAAGCCGATACAGCTGTGAC 360  
 QY 121 ArgLeuHisAspLeuThrThrGlyLeuSerAlaAlaThrPheValaAlaThrAsnGlyArg 140  
 Db 361 CGGCTCCAGCACTTGTGACTTTCGGAAGCTGGAAGCTGGACCAAGCTGCT 420  
 QY 141 ArgTTPArgThrGluArgGluMetTyrGlyCysAspGlyTyrValaArgAlaValaGluIle 160  
 Db 421 CGTTGGCGTAACGAACTGTAATGTACCGCTCGACGATACGTCGCGCTGTTGAAATC 480  
 QY 161 ThrHisGlyLysAsnGlyThrHisValaHisValaHisAlaLeuLeuMetPheSerGlyAsp 180  
 Db 481 ACTCAAGGAAAGAAAGCTGGGCACTGTCAGCTTCAAGGCTCACTCATGTCAGTGTGAC 540  
 QY 181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTTPThrSerLys 200  
 Db 541 GTGAGTGAAGAACATCTCTGGAATCTTCTCGATGCGATGCGATGCTGAGCTTCCAAA 600

QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgPheSerGlyGlyLeuAspValArgLys 220  
 Db 601 CTGATATCTCGGAGTTTGTGCGGCACTACGTAATTCGGGTGTCTCGAGTACGAAAG 660  
 QY 221 IleGlyGlyGluAlaAspGlnValLeuAlaAlaThrLeuThrLysIleAlaSerGlyVal 240  
 Db 661 ATCCGCGGTGAAGCTGACCAAGTCTCTCGCTCGATTCGACGAAATTTGATCTGGGCTT 720  
 QY 241 GlyMetGluValGlySerGlyLysArgGlyLysSerGlyArgHisGlyAsnArgAlaProTyr 260  
 Db 721 GGTATCGAGTTGTGATGCGGACGCGAAAGGTGTCAATGCGCAACCGTGCACCTCGG 780  
 QY 261 GluIleAlaValaAspAlaValaGlyLysAspProGlnAlaLeuGluLeuThrArgGluPhe 280  
 Db 781 GAAATCGCTGTGATGACGTGGCGGATCCGACATCAAGGTGTGAACCTGTGGCGAGAAATT 840  
 QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaThrPheSerArgGlyLeuArgAlaArgAla 300  
 Db 841 GATTTGGTTGATGCGGACGTGGCGGATCCGATCGGATGCTCCGATGATTCGTCGCCAGCT 900  
 QY 301 GlyLeuGlyValaGluLeuThrAspAlaGlnIleValaGlnGlnGlnLysSerAlaProVal 320  
 Db 901 GGTCTTGGCGGAACTAACAAGATGCTCAATCTGTGACAGAAATTCGCCCGGCTC 960  
 QY 321 MetValaAlaIleIleProAlaArgSerTrpMetIleArgThrCysAlaProTyrVal 340  
 Db 961 ATGTTGGATATATTCGGGCGGATCGGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 QY 341 PheGlyGluIleLeuGlyLeuValaGluAlaGlyValaThrTTPGluAsnLeuArgAspHis 360  
 Db 1021 TTCGGGAGATCTCGGACTCGTCCAACTGCGCGGACCTTGGGAAATCTTCGATGATCAC 1080  
 QY 361 LeuHisTyrArgLeuProAlaAlaAspValaArgProProIleIleSerValArgLys 379  
 Db 1081 TTGCATTAATGATTCGCCGCGGACGAGATGTGCGGCCCCGAAATATTCGTTGCAAG 1137

RESULT 2  
 AY178757/c 6334 bp DNA linear BCT 01-MAR-2003  
 LOCUS Rhodococcus erythropolis cryptic plasmid pAN12, complete sequence.  
 DEFINITION AY178757  
 ACCESSION AY178757  
 VERSION AY178757.1 GI:28628256  
 KEYWORDS  
 SOURCE Rhodococcus erythropolis  
 ORGANISM Rhodococcus erythropolis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Nocardiaceae; Rhodococcus.  
 REFERENCE  
 1 Kostichka, K., Tao, L., Brummel, M., Tomb, J.-F., Nagarajan, V. and  
 Cheng, Q.  
 A small cryptic plasmid from Rhodococcus erythropolis:  
 Characterization and utility for gene expression  
 Unpublished  
 Direct Submission  
 Submitted (11-NOV-2002) CRD, E.I. Dupont de Nemours Inc.,  
 Experimental Station, Wilmington, DE 19880-0328, USA  
 TITLE JOURNAL  
 AUTHORS Kostichka, K., Tao, L., Brummel, M., Tomb, J.-F., Nagarajan, V. and  
 Cheng, Q.  
 JOURNAL Submitted (11-NOV-2002) CRD, E.I. Dupont de Nemours Inc.,  
 Experimental Station, Wilmington, DE 19880-0328, USA  
 FEATURES  
 source 1..6334  
 /organism="Rhodococcus erythropolis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1833"  
 /plasmid="pAN12"  
 /note="cryptic plasmid"  
 /complement(1912..3051)  
 /gene="rep"  
 /complement(1912..3051)  
 /gene="rep"  
 /note="Rep"  
 /codon\_start=1



```

/translation=1
/product="replication protein"
/db_xref="GI:28628257"
/translation="MTSVAEHLGKDPDPVVSDDKRGIRHLEAPKIQITTSFTN
ACGRPISSGVNVTIVANGPGSGFGLRSCKGKMI CPCCKGKRGARADEISQVVAHOL
GTGSAVATWMTHTAGORLHDLMTGLSAAKMAITGRKMTFEREMADGDDYRAVEI
THKNGHVAHVHALLMSPGVSNTLESNDAMPDFTWSKLSLGPAPLRNSGLDV
RKIGSADVDLAIVLTIKASGVMEVSGSDGSGKGNAPMEIIVDAVCGDQALDEL
KREREGSMRRRIAMSRLRPAAGIATDQIVQESAPVWVAITIPASMMITR
TCAPVFEGLIGVEAGATWENDRHLYRLPADVRPPTISVRK"
complement(4288..5178)
/gene="div"
/complement(4288..5178)
/gene="div"
/feature="div"
/codon_start=1
/translation=1
/product="putative cell division protein"
/db_xref="GI:28628258"
/translation="MDQDTPIALIGMNLAPWVVDIAKDAHVLIOGKRSKSGOC
TYLLAAGSNPAVRYVGVDPSTSVLAPVHRPAPNTELENDPDKYLRLOFVKA
ESDRRIKCEMDRITDKISLSPALPILVLESPFQILEGADFPATNGLEKPADRYAP
RITSLVQIAAGAKAGIEMLLAQRASIVGNAKSFVAKMTLRVDEPESVRLH
PNATPECALVEGFPVGGFDPGLRQIMRTVRVGEYSIYASVENADLAYEALN
IDRAQMTTISEPRLGDIQ"
BASE COUNT 1386 a 1807 c 1681 g 1460 t
ORIGIN
Alignment Scores:
Pred. No.: 5.68e-118 Length: 6334
Score: 2005.00 Matches: 379
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-007-527a-2 (1-379) x AY178757 (1-6334)
QY 1 MetThSerValSerIaGluHISleuSerGIyLysAspArgProProValIleuValSer 20
Db 3051 ATGACCAAGCTTAAGTGTGTAACACCTTTCCGCAAGACCGGCTCCCTCGTGTGTCG 2992
QY 21 SerAspLysArgIyIleArgHISgluLeuArgProLysLeuGlnIleThrThrSer 40
Db 2991 TCCGATTAAGCGCGCATCCGCAAGAACTGCAACCAACTCAACAAATCACACGTCAT 2932
QY 41 GluThrPheAsnIaCysGlyArgProIleSerGIyValAsnGIyValThrIleValAsn 60
Db 2931 GAAACATTTAACCCCTGTGGCCGCGCATTTCTGCGTGAACCGTGTGACCAATGTCAAC 2872
QY 61 GlYProLysGIySerGIyPheGIyLysArgSerCysGIyLysGIyTrpIleCysPro 80
Db 2871 GGTCCGAAGGTTCTGATTCGAGAGGCTTCCTCTCGGGAAGGGCTGATCTGCCCC 2812
QY 81 CysCysAlaGIyLysValGIyAlaHISArgAlaAspGlnIleSerGlnValAlaHIS 100
Db 2811 TGGCTGCGGGAAGAGCGGTGCACTCGTGCAGAGAAATTTCTCAAGTGTGTGCTCAT 2752
QY 101 GluLeuGIyThrgIySerValIaMetValThrMetThrMetArgHISThrIaGlyGln 120
Db 2751 GAACTCGGAGCTGGATCTGTGGATGGTACATGACATGCGGCATACAGCTGGTCAG 2692
QY 121 ArgIleuHISAspLeuTrpHrgIyLeuSerAlaIaTrpLysAlaIaThrAsnGIyArg 140
Db 2691 CGGCTCCAGCACTATGACCTGACCTTCGCGACCTCGGAAAGCTCGACCAACGAGTCGT 2632
QY 141 ArgTrpArgHrgIyuaArgIyMetTrpGIyCysAsnGIyTrpAlaArgAlaIaGluIle 160
Db 2631 CGTGGCGTACGGAACCTGAATGTACGCTGCGACGGAATACCTGGCGCTGTGTAATC 2572
QY 161 ThrHISGIyLysAsnGIyTrpHISValHISValHISAlaLeuMetPheSerGIyAsp 180

```

```

Db 2571 ACTTACCGAAGAAACCGCTGGACGCTCCACGTTCAAGCCGCTACTATGTTCACTGTGAC 2512
QY 181 ValSerGIuAsnIleLeuGIuSerPheSerAspAlaMetPheAspArgTrpThrSerLys 200
Db 2511 GTGAGTAGAAGAACCTCCGAAATCCTTCTCGAAGGACATGTCATCGGTGACCTTCAAA 2452
QY 201 LeuValSerLeuGIyPheAlaIaPoluArgAsnSerGIyGIyLeuAspValaArgLys 220
Db 2451 CTGCTATCTCTGGATTTGCTGCGCCACTACGTAAATTCGGGTGCTCCATGTACAAAG 2392
QY 221 IleGIyGIyLysIaAspGlnValIleuAlaIaTrpLeuThrLysIleAlaSerGIyVal 240
Db 2391 ATCGCGCGTGAAGCTGATTCAGATTTCTCCCTCGCTATCTCAAGAAATTCATGCGCTT 2332
QY 241 GlYMetGIyValGIySerGIyAspGIyLysSerGIyArgHISGIyAsnArgAlaProTrp 260
Db 2331 GGTATGAGGTTGGTATGCGCGACGGAAGAGGTGTGACATGCAACCGTGCACCTCG 2272
QY 261 GluIleAlaValaAspAlaValaGIyGIyAspProGIaIaLeuGIuLeuTrpArgGluPhe 280
Db 2271 GAATTCCTCTGTTGATGCGAGTGGCGGGGATCCCAACGCTTGGAATCTGTGGCAGAAAT 2212
QY 281 GluPheGIySerMetGIyArgArgAlaIaIaTrpSerArgGIyLeuArgAlaArgAla 300
Db 2211 GAGTTTGTTCATGATGGACGCTCGGCAATCGCGGTGTCCTCGATTTGCTGCCAGCT 2152
QY 301 GlYLeuGIyLysIleuLeuThraAspAlaGlnIleValGIuGlnIleGlnIleuSerAlaProVal 320
Db 2151 GGTCTGGGCGAAGCTTACAGATGCTCAATGCTGTGAGAGAGAAAGATTCGCCGCTC 2092
QY 321 MetValaIaIleIleProAlaArgSerTrpMetMetIleArgTrpCysAlaProTrpVal 340
Db 2091 ATGTTTCGATCATTCCTCGGCGCATGCTGATGATGATGATGATGATGATGATGATGATG 2032
QY 341 PheGIyGIyLysLeuGIyLeuValGIyAlaGIyAlaIaThrTrpGIuAsnLeuArgAspHIS 360
Db 2031 TTGCGCAGATCTCCGCACTGTCGAAAGCTGGCGCGCATTCGGGAAATCTTCGTGATCAC 1972
QY 361 LeuHISTrpArgLeuProAlaIaAspValaArgProProIleSerValaArgLys 379
Db 1971 TTGCAATTATCATTTGCCCGCAGCGAGTGTGCGGCCCGCATATATGATGCTTCGACAG 1915
RESULT 3
AX548646/c 6334 bp DNA linear PAT 27-NOV-2002
LOCUS AX548646
DEFINITION Sequence 5 from Patent WO0205709.
ACCESSION AX548646
VERSION AX548646.1 GI:25813614
KEYWORDS
SOURCE
ORGANISM Rhodococcus erythropolis
Rhodococcus erythropolis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE 1. Brumucci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.
TITLE Rhodococcus cloning and expression vectors
JOURNAL Patent: WO 0205709-A 5 18-JUL-2002;
E.I. DU PONT DE NEMOURS AND COMPANY (US)
FEATURES
source 1..6334
/organism="Rhodococcus erythropolis"
/mol_type="genomic DNA"
/strain="AN12"
/db_xref="taxon:1833"
BASE COUNT 1386 a 1807 c 1681 g 1460 t
ORIGIN
Alignment Scores:
Pred. No.: 5.68e-118 Length: 6334
Score: 2005.00 Matches: 379
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0  
 US-10-007-527a-2 (1-379) x AX548646 (1-6334)

QY 1 MetThrSerValSerAlaGluHisLeuSerGlyValAspArgProProValLeuValSer 20  
 DB 3051 ATGACGAGCGTAAGTGTGAACACCTTCCGGCAAGACGGCTCCCGCTCGGTGTG 2992

QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSer 40  
 DB 2991 TCCGATTAAGCCGCGACATCCGACGAACTGCGACCAACTTCAACAAATCCACAGTGA 2932

QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60  
 DB 2931 GAACAATTAAACGCTGTGCGCCGCGCATTTCTGGCGTGAACGCTGTGACCTGTGCAAC 2872

QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrPheCysPro 80  
 DB 2871 GGTCCGAAGGTTCTGGATTCCGAGGCGCTTCTGCTGCGAAGGCGTGGATCTGCCCC 2812

QY 81 CysCysAlaGlyLysValGlyValHisArgAlaAspGluIleSerGlnValValAlaHis 100  
 DB 2811 TGTCTGGCGGAAAAGTCGGTGCACATGTCGACAGAAATTTCTCAAGTTGTGTCTCAT 2752

QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120  
 DB 2751 CAACCTCGGAGCTGGATCTGTTCGATGTGTGACATGACATGCGCATACAGCTGGTCTAG 2692

QY 121 ArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrPheValAlaThrAsnGlyArg 140  
 DB 2691 CGGCTCCACACCTATGAGCTGAGCTTTCCGACGCTGGAAAGCTCGACCAACGCTGCT 2632

QY 141 ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160  
 DB 2631 CGTTGGCGTACGGAACCTGAATGTACGCGCTGCGACGCAATCGCGCCCTGTGTAATC 2572

QY 161 ThrHisGlyLysAsnGlyTyrPheIleValHisValHisAlaLeuLeuMetPheSerGlyAsp 180  
 DB 2571 ACTCCAGAAAACCGCTGCGACGCTCCACGTTCCACCGCTCATCTCATGTTCATGTGTGAC 2512

QY 181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTyrThrSerLys 200  
 DB 2511 GTGATGAGAACATCTCTCGAATCTTCTCGATGCGATGTTCATGCGTGGACTTCCAA 2452

QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLys 220  
 DB 2451 CTGCTATCTCGGATTTGCTGGCGCACCTACGTAAATTCGGGTGTGTCTCGATGACAAAG 2392

QY 221 IleGlyGlyGluAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyVal 240  
 DB 2391 ATCGCGGTGAGAGCTGATCAAGTCTCGTGCATGTGACGAAATTTGATCTGGCGAT 2332

QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyValArgHisGlyAsnArgAlaProTyr 260  
 DB 2331 GGTATGAGAGTGTGTATGTGGCGACGAAAAGTGTGTGACATGGCAACCTGTGACCTCG 2272

QY 261 GluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeuGluLeuTyrPheGluPhe 280  
 DB 2271 GAAATCGCTGTGTGAGTGGGCGGGGATCCACAACTGTGGAACCTGTGCGACAGAAATT 2212

QY 281 GluIlePheGlySerMetGlyArgArgAlaIleAlaThrSerArgGlyLeuArgAlaArgAla 300  
 DB 2211 GAGTTGGTTCGATGGGACGTCGGGCAATGCGGTGGTCCGTTGGATTCGTCGCCAGCT 2152

QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluGluSerAlaProVal 320  
 DB 2151 GGTCTGGGCGGACGAATCAAGATGCTCAAGTCTTGAACAGAGAAATCTGCCCGCGTC 2092

QY 321 MetValAlaIleIleProAlaArgSerTyrMetMetIleArgThrCysAlaProTyrVal 340  
 DB 2091 ATGATTCGATCATTCGGCGCGCATGTGATGATGATTCGGAATCTGTGTGCGCTTACGTC 2032

QY 341 PheGlyGluIleLeuGlyLeuValGluAlaGlyValThrTyrPheGluAsnLeuArgAspHis 360  
 DB 2031 TTCCGCGAGATCTCTCGACCTGCTCGAAGCTGGCGCACCTTGGAAAATCTTCTGATCAC 1972

QY 361 LeuHisTyrArgLeuProAlaAlaAspValArgProProIleLeuSerValArgLys 379  
 DB 1971 TTGCATTAATGATTCGCCGCGACGCGATGTGTGCGCCCCCGATTAATATGCTTCCAG 1915

RESULT 4  
 AX548648/c  
 LOCUS AX548648 9652 bp DNA linear PAT 27-NOV-2002  
 DEFINITION Sequence 7 from Patent WO2055709.  
 ACCESSION AX548648  
 VERSION AX548648.1 GI:25813616  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Shuttle vector pRHR17  
 Shuttle vector pRHR17  
 artificial sequences; vectors.  
 REFERENCES  
 1. Brumaci, M.G., Cheng, Q., Kostichka, K.N. and Tomb, J.F.  
 "Rapid cloning and expression vectors  
 Patent: WO 02055709-A 7 18-JUL-2002;  
 E.I. DU PONT DE NEMOURS AND COMPANY (US)  
 FEATURES  
 source  
 1. 9652  
 /organism="Shuttle vector pRHR17"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:214938"  
 BASE COUNT 2097 a 2729 c 2559 g 2267 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 9,26e-118 Length: 9652  
 Score: 2005.00 Matches: 379  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-007-527a-2 (1-379) x AX548646 (1-9652)

QY 1 MetThrSerValSerAlaGluHisLeuSerGlyValAspArgProProValLeuValSer 20  
 DB 5052 ATGACGAGCGTAAGTGTGAACACCTTCCGGCAAGACGGCTCCCGCTCGGTGTG 4993

QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSer 40  
 DB 4992 TCCGATTAAGCCGCGCATCCGCGACGAGACTGGAACCAACTTCAACAAATCACACGTC 4933

QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60  
 DB 4932 GAACAATTAAACGCTGTGCGCCGCGCGCATTTCTGACGTGAACGCTGTGACCATTTGTC 4873

QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrPheCysPro 80  
 DB 4872 GGTCCGAAGGTTCTGGATTGGAAGGCTTCTGCTGCGGAAAAGGCTGGATCTGCCCC 4813

QY 81 CysCysAlaGlyLysValGlyValHisArgAlaAspGluIleSerGlnValValAlaHis 100  
 DB 4812 TGTCTGGCGGAAAAGTCGGTGCACATGTCGACACCAAAATTTCTCAAGTGTGTCTCAT 4753

QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120  
 DB 4752 CAACCTCGGAGCTGGATCTGTTCGATGTGACATGACATGACATGACATGACATGACAT 4693

QY 121 ArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrPheValAlaThrAsnGlyArg 140  
 DB 4692 CGGCTCCACGACCTATGACTGTGACTTTCGCGACCTGGAAGAGCTGGACCAACGATCGT 4633

QY 141 ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160  
 DB 4632 GGTGTGCGTAAGAAACGTAATATACGCTGCGACGAGATGATGCGCGCTGTGGAATC 4573



Db 5621 TTGGCGAGATCTCGGACTCGTCCGAGCGACTTGGGAAAATCTCGTATCAG 5562  
 Qy 361 LeuHISYrArgLeuProAlaAlaAspValArgProProIleIleSerValArgIys 379  
 Db 5561 TTGCATTATGATTGCGCCGAGCGAGTGTGGGCCCCCGCATATATCGATTGCAAG 5505

RESULT 6  
 AY180162/c 11241 bp DNA circular SYN 01-MAY-2003  
 LOCUS Shuttle vector pRHBRI7, complete sequence.  
 DEFINITION AY180162  
 ACCESSION AY180162  
 VERSION AY180162.1 GI:30313707  
 KEYWORDS  
 SOURCE Shuttle vector pRHBRI7  
 ORGANISM Shuttle vector pRHBRI7  
 REFERENCE 1 (bases 1 to 11241)  
 AUTHORS Kostichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and Cheng, Q.  
 TITLE A small cryptic plasmid from Rhodococcus erythropolis: characterization and utility for gene expression

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 11241)  
 AUTHORS Kostichka, K., Tao, L., Brammucci, M., Tomb, J.-F., Nagarajan, V. and Cheng, Q.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-NOV-2002) CR@P, E. I. Dupont de Nemours Inc., Experimental Station, Wilmington, DE 19880-0328, USA  
 FEATURES  
 source  
 1..11241  
 /organism="Shuttle vector pRHBRI7"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:214938"  
 /focus  
 /note="Rhodococcus - Escherichia coli shuttle vector: Ssp I linearize PAN12 cloned into the PvuII site of pBR328"  
 join(9925..11241,1..3590)  
 /organism="Cloning vector pBR328"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:221946"  
 3591..9924  
 /organism="Rhodococcus erythropolis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1833"  
 /plasmid="PAN12"  
 /note="cryptic plasmid"

BASE COUNT 2517 a 3109 c 2945 g 2670 t  
 ORIGIN

Alignment Scores:  
 Pred. NO.: 1.1e-117 Length: 11241  
 Score: 2005.00 Matches: 379  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 12 Gaps: 0

US-10-007-527a-2 (1-379) x AY180162 (1-11241)

Qy 1 MetHrSerValSerAlaGluHisLeuSerGlyLysAspArgProProValIleuValSer 20  
 Db 6641 ATGACCAAGCGTAACTGTGAACACCTTCCGCAAGAACCGGCTCCCGTCCGCTGTG 6582  
 Qy 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40  
 Db 6581 TCCCATTAAGCGCGCATCCGCAACGAACTCGAACCCCAAACTTCACAAATCACACAGTCA 6522  
 Qy 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60  
 Db 6521 GAACATTTAAACGCTGTGCGCGCGCATTTCTGGCGTGAACGCTGTGACCATTTGCAAC 6462  
 Qy 61 GlyProLysGlySerGlyPheGlyLysLeuArgSerGlyLysGlyTrrIleCysPro 80

Db 6461 GGTCCGAAAGGTTCTGGAATTCGAGGCGTTCCTCGGAAAAGGCGTGAATCTGCCCC 6402

Qy 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValValAlaHis 100  
 Db 6401 TGCCTGGGAGAAAAGCGGTGACATCGTCGAGAGAAATTTCTCAATTTGTGCTCAT 6342

Qy 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyIn 120  
 Db 6341 CAATCGGAGCTGATGATGTTGCGATGATGATGATGATGATGATGATGATGATGATGAT 6282

Qy 121 ArgLeuHisAspLeuTrpThrGlyLeuSerAlaIleTrpValAlaIleThrAsnGlyArg 140  
 Db 6281 CGGCTCCACCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6222

Qy 141 ArgTrpArgThrGluArgGluMetTrpGlyCysAspGlyTrpValArgAlaGluIle 160  
 Db 6221 CGTTGGGTCGAGCACTGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 6162

Qy 161 ThrHisGlyLysAsnGlyTrpHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180  
 Db 6161 ACTCAGCAAAAAACGCGCTGCAAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAG 6102

Qy 181 ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTrpThrSerIys 200  
 Db 6101 GTGATGAGAACATCTCGATGATCTCTCGATGATGATGATGATGATGATGATGATGATGAT 6042

Qy 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysLeuAspValArgIys 220  
 Db 6041 CTGCTATCTCTGGATTGCTGCGCCACTGATGATGATGATGATGATGATGATGATGATGAT 5982

Qy 221 IleGlyGlyGluAlaAspGlnValLeuAlaIleTrpLeuThrLysIleAlaSerGlyVal 240  
 Db 5981 ATCCGCGGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGTAAAGTGT 5922

Qy 241 GlnMetGluValGlySerGlyLysAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrp 260  
 Db 5921 GGTATGAGGTTGTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5862

Qy 261 GluIleAlaValAspAlaValGlyLysProGlnAlaLeuGluLeuTrpArgGluPhe 280  
 Db 5861 GAATTCCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5802

Qy 281 GluPheGlySerMetGlyArgArgAlaIleAlaIleTrpSerArgGlyLeuAlaAlaGlyAla 300  
 Db 5801 GAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5742

Qy 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGlnGlnSerAlaProVal 320  
 Db 5741 GGTCTGGGCGAGAACCTAACAGATGCTCAATGCTTGAAGAGAGAGATGCTGCCGCTC 5682

Qy 321 MetValAlaIleIlePheAlaArgSerTrpMetMetIleArgThrCysAlaProTrpVal 340  
 Db 5681 ATGTTGGAGCATTTCCGCGCGCATCGTGAATGATGATGATGATGATGATGATGATGATGAT 5622

Qy 341 PheGlyGluIleLeuGlyLeuValGluAlaGlyValAlaThrTrpGluAsnLeuArgAspHis 360  
 Db 5621 TTCCGCGAGATCTCGAATCTGTCGAAAGCTGGCGCACTTGGGAAAATCTTCGTGATCAC 5562

Qy 361 LeuHISYrArgLeuProAlaAlaAspValArgProProIleIleSerValArgIys 379  
 Db 5561 TTGCATTATGATTGCGCCGAGCGAGTGTGGGCCCCCGCATATATCGATTGCAAG 5505

RESULT 7  
 AP083788 2439 bp DNA circular BCT 15-JUN-1998  
 LOCUS Actinomyces pyogenes plasmid pAP1, complete plasmid sequence.  
 DEFINITION U83788  
 ACCESSION U83788  
 VERSION U83788.1 GI:1805288  
 KEYWORDS  
 SOURCE Arcanobacterium pyogenes  
 ORGANISM Arcanobacterium pyogenes  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE  
1 (bases 1 to 2439)  
Billington, S.J., Jost, B.H. and Songer, J.G.  
The Arcanobacterium (Actinomyces) pyogenes plasmid papi is a member  
of the pU101/pU1 family of rolling circle replication plasmids  
U. Bacteriol. 180 (12), 3233-3236 (1998)

JOURNAL  
MEDLINE  
96292760  
PUBMED  
9620977  
2 (bases 1 to 2439)  
Billington, S.J., Jost, B.H. and Songer, J.G.  
Direct Submission  
Submitted (06-JAN-1997) Veterinary Science, University of Arizona,  
117 East Lowell Street, Tucson, AZ 85721, USA  
Location/Qualifiers

FEATURES  
source  
1. .2439  
/organism="Arcanobacterium pyogenes"  
/mol\_type="genomic DNA"  
/strain="BB1"  
/db\_xref="taxon:1661"  
/plasmid="papi"  
1. .110  
/note="single stranded origin; putative"  
39. .44  
/note="size of single strand initiation; putative"  
221. .1600  
/gene="rep"  
221. .1600  
/gene="rep"  
/function="replication protein"  
/codon\_start=1  
/transl\_table=11  
/product="Rep"  
/protein\_id="AAC46399.1"  
/db\_xref="GI:1805289"  
/translation="MNRISRTALSLPAQIQKVPAPAGSLKSLFEGMTATWSAGG  
ASSSRRDRKSOIPSNRRRGSAATHPLGNTVLPVNSKKTAKRSERYELDQ  
LAESTIESVRKGRVPAPVAPVSLRAKSDKAGVGLTGGVWACPRCSAKIAR  
KTDIQVVDHVAKGMTVSMVTLQRHKGQGLKHLMDLSTAMNVRVIGRENTKE  
QFGLVGVANAEITHGQGWVSHVLI1SEKDLSTFYORRQGRRLPPPELTM  
SDPFIARWRAEALAKRGVDRDGGIDMTVAADARAIGNVYRKQTSDAISSEVTL  
GPFKAKNRPPTPOLADILSLGDVDDLTKREYKASFGRAALTWKGLDMANIG  
VEQSDDEIASEBRIQDEALFLTHDAMQVRRFGAELLDVTEGGRALVYRMIDPREI  
DMSLPKIE"

## RBS

CDS

1673. .1676

1685. .2023

/note="ORF112"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAC46400.1"

/db\_xref="GI:1805290"

/translation="MFTIGKRTKFPQPIADCKSPFPDNGKRTGSSSEKMTQAIT  
EDGLVBLTSPFQFQVLDVAKSRLEFEKVNLAFLPRNGGISYVLTGEAVHV  
PAPVAVGVSE"

RBS

CDS

2010. .2013

2020. .2409

/note="ORF129"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAC46401.1"

/db\_xref="GI:1805291"

/translation="MMSBPYGISNNAFLQVSPRPAKRAEATYMSRIKKELSGHGS  
TPTALASCEKLTAVSEPRIVDVVHLHEVGECCACAGTNSAVASLSRAFTIPARFQS  
LENGRNPAQLLTQDVAGMKTLGT"

rep\_origin

BASE COUNT

603 a

657 c

656 g

523 t

/note="double stranded origin; putative"

ORIGIN

603 a

657 c

656 g

523 t

/note="double stranded origin; putative"

Alignment Scores:  
Pred. No.:  
Score:

7.77e-30

609.50

Length:

Matches:

2439

148

Percent Similarity: 51.47%  
Best Local Similarity: 36.27%  
Query Match: 30.40%  
DB: 1  
Gaps: 15

US-10-007-527a-2 (1-379) x AP083788 (1-2439)

QY 5 SerIaGIuHISLeuSerGIYyAspArgProProValLeuValSerSerAsp----- 22  
DB 419 TCAGAGACCATCCCTTGGCAATAGCGATTAACTTTCAGATCAAGAGAGAGAG 478  
QY 23 -----LysArgIYLlAArgHISLeuLeuArgProLysLeuGlnIle 37  
DB 479 AAAACAGCTAATCTCGCGCTTCTGAGAGATACCACTCAGAGCGATTAGCCGAATC 538  
QY 38 ThrThiSerGIuHrPheAsnAlaCysGIYArg---ProIleSerGIYValAsnGIYVal 56  
DB 539 TCAGCATTTGATTCGTCGCGAAGTGGCCGCGTCCGCGGACCTTCGTCGTTG 598  
QY 57 ThrIleValAsnGIYProLysGIYSerGIYpHeGIYLeuArgSerGIYLyGIY 76  
DB 599 CGAGCAAAATCTGACGGTAAAGCGCGGATATGTTGTTGACACCTTGTGAAGGCTC 658  
QY 77 TrpIleCysProCysCysAlaGIYLyValGIYValAsnArgAlaAspGIYLeuSerGIY 96  
DB 659 TGGGCTCCCACTGCTGTAGCCGAAATCGCCCTCGCGGAAAAACCGACTTCACACG 718  
QY 97 ValValAlaHISGlnLeuGIYThrGIY---SerValAlaMetValThrMetThrMetArg 115  
DB 719 GTTCGTGACCAAGCGGTAAACAGAGATGACGCTTCATGCTTACGCTCACCAAGCT 778  
QY 116 HisThrAlaGIYValAsnLeuHisAspLeuTrpThrGIYLeuSerAlaAlaTrpLyAla 135  
DB 779 CACCAAGAGAGCAAGGCTAAACACCTTCGAGACGCTTGTGAGCGCTGGAATCGC 838  
QY 136 AlaThrAsnGIYArgArgTrpArgThrGIYArgIuMetTrpGIYCysAspGIYTrpVal 155  
DB 839 GTTACCTTGTGCTGCTGTTGATTTGATTCAGAGCAATTTGTTAGTGTGTTAGT 898  
QY 156 ArgAlaValGIYLeuThrHISGIYLyAsnGIYTrpHISValHISValHISAlaLeuLeu 175  
DB 899 CGAGCAATGAATTAATCATGAGAAAGCAGCGTGGATGATTCATCCATGTTCTGATT 958  
QY 176 MetPheSerGIYAspValSerGIYAsnIleLeuIuSerPhe----- 189  
DB 959 ATTTCCGAGAAAGAC-----CCGCTGACAGACGTTGTCTATCAACGAAACAA 1009  
QY 190 -----SerAspAlaMetPheAsp 195  
DB 1010 GAGAGCGCGCGCTTCCCTACCCCGCAGAGATTATATGTCATCGATTTCATGCTGAA 1069  
QY 136 ArgTrpThrSerLyLeuValSerLeuGIYpHeAlaAlaProLeuArgAsnSerGIY 215  
DB 1070 CGGTGGAAAGCTGGCTTGCAGAGACGCGCTTGAATTT---CTCCCGGATTCGAGGCG 1126  
QY 216 Leu-----AspValArgHISLeuGIYValAlaAspGlnValLeu 229  
DB 1127 TTGACGTGACCGTTGCGAAAGAGCGCGGACCATTCGCAAC----- 1168  
QY 230 AlaAlaIYLeuThrLyIle-----AlaSerGIYValGIYMetGIYValGIYSer 246  
DB 1169 -----ThrGTGAGAAATGCGACGTCACAGAGGATTAAGTCGAGAGTCAAGTTG 1222  
QY 247 GIYAspGIYLySerGIYArgHISGIYAsnArgAlaProTrpGIYValAlaValAspAla 266  
DB 1223 GCGGCGTTCAAAAGAGCGCGAAAGCGAGAGCGCTTCAGATACCTCGCGGATATC 1282  
QY 267 ValGIY---GIYAspProGIYAlaLeuGIYLeuTrpArgIuPheGIYSerMet 285  
DB 1283 CTTTGGCTCGGAGATGTGAGAGACCTCAAGCTCTGAGAAAGATATGAGAAAGCTTCGTT 1342  
QY 286 GIYArgArgAlaIleAlaTrpSerArgGIYLeuArgAlaArgAlaGIYLeuGIYAlaGIY 305



gene	1979..3280
CDS	/gene="rep" 1979..3280 /gene="rep" /note="plasmid replication protein" /codon_start=1 /transl_table=1 /product="rep" /protein_id="AA078123.1" /db_xref="GI:27465057" /translation="MNIIVTISKQPHDLCSHEPCARRSERORAGDRLGTTTRA FPVSDSPKRNENGRRRHYEMRDGLRNPQWPLRLVRKCAVPQSORTALMAGGAG YAGATGSCVWACPVCAKISAHREDELARVQVAFGVSMLTLLORHAGGDLLE ELMSTLOSQKMAVTSGRMOEFCAQLGVQGVKRVKVEYTHSGHGHVHTVYIKQDI TSNDLTIHRKKAQRRRTPYEPYEVQRPDPLAEMRSKLRKRGDFLAGSGGLDQIT DSGBEALGRYVAKMNSVDGLANBATTUGFKKARNGRTFQILDELDTGSEETDL LMRYVAFSAHGKALTWKSLRMDWAGMESMSDQVAADQCGEAVLFDHDAMROIIT TAGAAFLDLELHSGSVVAMLKRRIRHYEIPVPMWSTP"
BASE COUNT	693 a 967 c 1066 g 814 t
ORIGIN	
Alignment Scores:	
Pred. No.:	4.46e-27
Score:	569.00
Percent Similarity:	51.78%
Best Local Similarity:	35.79%
Query Match:	28.38%
DB:	1 Gaps: 14
US-10-007-527A-2 (1-379) x AY150274 (1-3540)	
OY	4 ValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSerSerAspLys 23
Db	2094 GTACCAAGCGGACCACTTTCCGGGTATCCAGATTCCCTCTAGAA-MAAGCGAATGAGGCG 2155
OY	24 ArgGlyIleArgHisGluLeuArgProLysLeuGln-----GlnIleThrThrsSerGlu 41
Db	2153 CGTGGCCATGCTCTATGAGATGAGGAGATGCTTACGAATCCGCAAGTCATGCCGTGAG 2211
OY	42 ThrPheAsnAlaCysGlyArg--ProIleSerGlyValaGlnGlyValThrIleValAsn 60
Db	2213 CGCGTTCCGCAAGTCGGGGCAGTCCCGGTTTCG-----CAACGATCGCGTGAACGCG 2266
OY	61 GlyProLysGlySerGlyPheGlyGlyIleuArgSerCysGlyLysGlyTyrIleCysPro 80
Db	2267 GGTCAATGGTGGTCCGGTTATGCCGGTGTGGCGACGTCCGGAAGTGTGTGGCGTTCCT 2322
OY	81 CysCysAlaGlyLysValaGlyAlaHisArgAlaAspGluIleSerGlnValaValaHis 100
Db	2327 GTCTGTGGCGGAAGAATTTCCGCCCAACCTGTGATAGCTGCGCCCGCTTGTCCAGGTT 2388
OY	101 GlnLeuGlyThrGly--SerValAlaMetValThrMetThrMetArgHisThrAlaGly 119
Db	2387 GCGGTGTGACACCGGCTTCAAGGTGTGATGTCGACGCTTACCAACGTCATCATGCTGGT 2446
OY	120 GlnArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrPylsAlaAlaThrAsnGly 139
Db	2447 CAGGATCTCGCCGAGCTGTGGCGCTCGCTCCAGTCGCGGTTGGAATGCTGTACACAGTGGT 2506
OY	140 ArgArgTyrPargThrGluArgGluMetCysGlyCysAspGlyTyrValaArgAlaValaGlu 159
Db	2507 CGAAGTGTGGCAGAAATTTTGGCGCTCAGCTCGCGCTCCAGAGATGGGTCAAGCACTTGA 2566
OY	160 IleThrHisGlyLysAsnGlyTyrPheIleValaHisValaHisAlaLeuLeuMetPheSerGly 179
Db	2567 GTCCACCATGAGTGTGATGGGTGTGACAGTTTCACTGACACGTCGTCTATCTTAAAGCAG 2622
OY	180 Asp-----ValSerGluAsnIle----- 185
Db	2627 GATCCGACATGACGTTGACACTAAGATTGGCAGTCGCCGAAACAAGTCGGCGCGGACCC 2686
OY	186 -----LeuGlnSerPheSerAspAlaMetPheAspArgTyrThrSerLys 200





QY 251 SerGIYArgHis-----GlyAsnArgAlaProTrrpGluIleAlaValaAspAlaValaGly 268  
 Db 597 ACGCGCGCTTCTCGCTCTCGACTCGATCGACTTGGAGGTCTTGGCGCGCGCGCAAC 538  
 QY 269 GlyAspProGlnAlaLeuGluIleuTrrpArgGluPheGlySerMetGIYArgArg 288  
 Db 537 GGTTCGGCCGAAGAAGGTGTCTCTGTTCGAGTGGAGAGGTTCGCGGAGCCAG 478  
 QY 289 AlaIleAlaTrrpSerArgGlyLeuArgAlaArgAlaGlyLeuGlyValaGluLeuThrAsp 308  
 Db 477 CAGATTGGCTGTCTCGCGGGTTGAGGAGACCGTCTCGGCTTCAAGTTGAGAAATCGAC 418  
 QY 309 AlaGlnIleValaGluGlnGlu---GluSerAlaProValMetValaIleIleProAla 327  
 Db 417 GACGTATGCGCGCTGAGAGAGTCCGGCTGTTCGACACTGTCCTTGTGATCAACGGGT 358  
 QY 328 ArgSerTrrpMet-----MetIleArgThrCysAla 337  
 Db 357 GACGATGGCGTCCGCTCGTGGGGGACAGAACCTTACAGAGTGTTCGAGAGTTC--- 301  
 QY 338 ProTrrpValPheGlyGluIleLeuGlyLeuValaGluAlaGly---AlaThrTrrpGluAsn 356  
 Db 300 -----GAATTGGCGCGT-----CAAGCTGGCTTGGCGCGGTGG----- 268  
 QY 357 LeuArgAspHis---LeuHisTrrpArgLeuProAlaAlaAsp 369  
 Db 267 CTGCAGAGACGAGATTCAGACACGATTGTGTGAGAGTGGAC 226

RESULT 10  
 SLMPG33A 2661 bp DNA linear BCT 09-MAY-1994  
 LOCUS S.lavendulae miniplasmid pSLG33 sequence.  
 DEFINITION X69872  
 VERSION X69872.1 GI:47179  
 KEYWORDS miniplasmid.  
 SOURCE Streptomyces lavendulae  
 ORGANISM Streptomyces lavendulae  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE  
 1 Felsberg, J., Petricek, M. and Richy, P.  
 Nucleotide sequence of the mini-plasmid pSLG33 from Streptomyces  
 lavendulae-grasserius RIA746  
 JOURNAL Nucleic Acids Res. 21 (15), 3582 (1993)  
 MEDLINE 8346038  
 PUBMED 93348001

REMARK  
 (sites)  
 2 (bases 1 to 2661)  
 TITLE Direct Submission  
 AUTHORS Submitted (28-DEC-1992) J. Felsberg, Inst. of Microbiology, Cell  
 and Molecular Microbiology Div., Videnska 1083, Prague 4,  
 CZBCHOSIOVAKIA

FEATURES  
 source location/Qualifiers  
 1..2661  
 /organism="Streptomyces lavendulae"  
 /mol\_type="genomic DNA"  
 /strain="RIA746"  
 /db\_xref="taxon:1914"  
 /dev\_stage="mycelia"  
 /plasmid="miniplasmid pSLG33"

BASE COUNT 514 a 916 c 818 g 413 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 7 45e-20 Length: 2661  
 Score: 453.00 Matches: 129  
 Percent Similarity: 48.33% Conservative: 59  
 Best Local Similarity: 33.16% Mismatches: 160  
 Query Match: 22.59% Indels: 41  
 DB: 1 Gaps: 16

US-10-007-527A-2 (1-379) x SLMPG33A (1-2661)

QY 3 SerValSerAlaGlnHisIleuSerGlyIysAspArgPro-----ProValLeu 18  
 Db 586 AATGTTCCCAACCCCTCCCTCTTCCCACTCAATCTCCAGCTCAGAGCCGATTCGCC 645  
 QY 19 ValSerSerAspIysArgGlyIleArgHisGluLeuArgProIysLeuGlnIleThr 38  
 Db 646 GTTTCGTCAAGGCGCCGGAGAACGCGCTTCGCCACGCGGAGTCTCTCGAGCATCACC 705  
 QY 39 ThrSerGluThrPheAsnAlaCysGlyIysArgProIleSerGlyValaAsnGly----- 55  
 Db 706 AGCTTCCCGCGCGTCCGAGTGCCTGGAGAACCAATGAGAGCGTCCGACCGCTGC 765  
 QY 56 ---ValThrIleValaAsnGlyProIysSerGlyPheGlyIleuArgSerCysGly 74  
 Db 766 GCGGAGATCGTGTGTAACGCG---TCCGTGCTCACTGTGTCCGCTTTCGACCTGGGT 822  
 QY 75 LysGlyTrrpIleCysProCysCysAlaGlyValaGlyAlaIleArgAlaAspGluIle 94  
 Db 823 TCCATCTGGGCTCGCCGCTGTCTCGGCAAGATCAGGGCCAAACCGCGGAGACGATC 882  
 QY 95 SerGlnValaAlaHisGlnLeuGlyThrGlySerValAla---MetValThrMetThr 113  
 Db 883 GCACAGCGCGCGCCAGACGCTCCGACGCGCGGACCGCTGGATGGTCACTGCTGAG 942  
 QY 114 MetArgHisThrAlaGlyGlnArgGluHisAspLeuTrrpThrGlyLeuSerAlaIleTrrp 133  
 Db 943 GCCCGACACAGAACACACACAGAGCTTACGCCCTCTTCGACGCGCTGCACACGGCTGG 1002  
 QY 134 LysAlaAlaThrAsnGlyIysArgTrrpArg-----ThrGlu 145  
 Db 1003 CGCAAGCTGTGTCCGCGAGGGCTTGGCCGAGAACCCCAAGCGGGAGTCTCCGCGCA 1062  
 QY 146 ArgGluMetTrrpGlyCysAspGlyTrrpValaArgAlaValaGluIleThrHisGly---Lys 164  
 Db 1063 CGCAGCGCGCTCGAGTCCGCGGCAACATCCGCTCACTCGAAGTCACTTACGGAATCGT 1122  
 QY 165 AsnGlyTrrpHisValaHisValaHisAlaLeuMetPheSerGlyAspValSerGluAsn 184  
 Db 1123 AACGTTGGCACCCGCACTTCAAGTTCCTTCCTGCTGCAAGAACGACCAACGACAA 1182  
 QY 185 IleuGluSerPheSerAspAlaMetPheAspArg---TrrpThrIysIysValSer 203  
 Db 1183 CTC-----GCTACGCAATGACACCGAGTGGACAAACTGTGAGGCGCTGCTCAAGAA 1236  
 QY 204 LeuGlyPheAlaAlaProLeuArgAsnSerGlyIleLeuAspValaArgLysIleGlyGly 223  
 Db 1237 GCCGATTGGAACGACGACGACGAAACAC-----GGCATCACCTGTGTCAAAATGC- 1284  
 QY 224 GluAlaAspGlnValaIleuAlaAlaTrrpThrIysIleAlaSerGly---ValGlyMet 242  
 Db 1285 ACGAACCCGAGAAAGCGGAGATACATCCCAAGCTCAAGAGGAAAGTCCGGAAC 1344  
 QY 243 GluValaGlySerGlyAspGlyIysSerGlyIysArgHisGlyAsnArgAlaProTrrpGlu--- 261  
 Db 1345 GAATTCGCCGCGAGCGACATGAAGCGCGTGTGCTCGGAATCTGGCGCCGTTGGAATG 1404  
 QY 262 IleAlaValaAspAlaValaGlyIysAspProGlnAlaLeuGluLeuTrrpArgGluPheGly 281  
 Db 1405 CTGCAATACCTTCGCGGACACCGGTGACATGGCATGGCTCCGATTTGGCAGAGATACAG 1464  
 QY 282 PheGlySerMetGlyArgArgAlaIleAlaTrrpSerArgGlyLeuArgAlaArg---Ala 300  
 Db 1465 AAGGACAGCTTCGCGCTCGGCGCATACCTGGAGCGTGGCTTCGCGCAATCTCTC 1524  
 QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValaGluGlnGluSerAlaProVal 320  
 Db 1525 GGAACAGAAAGAGCTCACGACGAGAACTTCAGCTGAGAGATCGAGAGGAGACG 1584  
 QY 321 MetValaIleIleProAlaIysSerTrrpMetIleArgThrCysAlaProTrrpVal 340  
 Db 1585 TGG---GCATTCGCGCGCGAATCGCTCCAGACGATCGCGCT---GTCCCGGCTC 1638

QY 341 PheGlyGluIleuGlyLeuValGluAlaGly-AlaThrTrp----- 354  
DB 1639 CAGTCCGATCCCTCGACACCGCCGAGACCGCGCTTGGAGACTGTGACCTGCTGA 1698  
QY 355 -----GluAsnLeuArgAsp 359  
DB 1699 CTGCCTACCATCTTCTCTACGAGAT 1723  
RESULT 11  
SNA243257  
LOCUS SNA243257 9367 bp DNA circular BCT 02-DEC-1999  
DEFINITION Streptomyces natalensis plasmid pSNA1 including spda, kora, spdb2  
and rep genes.  
ACCESSION AJ243257  
VERSION AJ243257.1 GI:6523476  
KEYWORDS KORA protein; rep gene; replication protein; Spda protein; SPDB2  
protein; transfer protein.  
SOURCE Streptomyces natalensis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Mendes, M.V., Aparicio, J.F. and Martin, J.F.  
TITLE Complete nucleotide sequence and characterization of pSNA1 from the  
plasmid-producing Streptomyces natalensis that replicates by a  
rolling circle mechanism  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 9367)  
AUTHORS Mendes, M.V.  
TITLE Direct Submision  
JOURNAL Submitted (23-JUN-1999) Mendes M.V., Microbiology, Inbiotec, AV.  
REAL 1, Leon, 24006, SPAIN  
FEATURES  
source  
1. 9367  
/organism="Streptomyces natalensis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:68242"  
/plasmid="pSNA1"  
/country="South Africa"  
266..814  
/gene="spda"  
266..814  
/gene="spda"  
/function="putative intramycelial transfer protein"  
/note="ORF4"  
/codon\_start=1  
/transl\_table=11  
/product="SPDA protein"  
/protein\_id="CAB62258.1"  
/db\_xref="GI:6523477"  
/translation="MTVIAHPHTGMPSPAGTVTGTGGEIPRDADGRRLCEHCLVWP  
VPSIGTKPRYSCSRNCQRAYEARQGFVQVTVMALLRSMRANKATQATSRNP  
EAKSRDADQATSRDVAAPATATPTAPABRPVPVQPTQALFEEDQADEBAP  
RPARMTDLKOMPRLPGFG"  
complement (1203..1940)  
/gene="kora"  
complement (1203..1940)  
/gene="kora"  
/function="putative transcriptional regulator"  
/note="ORF1"  
/codon\_start=1  
/transl\_table=11  
/product="KORA protein"  
/protein\_id="CAB62259.1"  
/db\_xref="GI:6523478"  
/translation="MPKIRASAPPYMVOVTHFRTRIMSGELSPGALPTDRGLAENG  
VSRATKAKALTVLKTGELIAGVSGTRQSGAPLPHNHNDPAVSVRGRIRYAGGEY  
ARITASLEPAPADYBAIGIEGAPALRRITVTNASVDALSMSTSWMDGLSTAP  
ALVSTRIRREGTWRVLETKGVVVGQDQSVRLATLEAQLLSLEPPAAVKLTRTI  
LRDEGILLEYSVTAGRESIYEYTV"  
2096..3025  
/gene="spdb2"  
2096..3025  
gene  
CDS

/gene="spdb2"  
/function="putative intramycelial transfer protein"  
/note="ORF2"  
/codon\_start=1  
/transl\_table=11  
/product="SPDB2 protein"  
/protein\_id="CAB62260.1"  
/db\_xref="GI:6523479"  
/translation="MPVRPVHRPPAGADPAQLERILRGLTVALLVVAAGLAFTATNV  
TOPADHGVSRIAMLDPLVASAGTVLVGGLSEFEVRPGASILRMFAGLATM  
LNCWGLSIPGTPGPIPTHLDPAGILHSULPYLLVLAITTRIRISIAVID  
EDTQGTPTDTPMPAPASAPVPSPTILASPPAPAPVPRORLEVCAGEOVR  
PALPARVSPKPEBPAPVPMGADPVALVPAWMDLSTROTARVGLSKTTVA  
NIMRELAHSGORPAGQPLPAGQTVLGVA"  
6656..8086  
/gene="rep"  
6656..8086  
/gene="rep"  
/note="ORF3"  
/codon\_start=1  
/transl\_table=11  
/product="replication protein"  
/protein\_id="CAB62261.1"  
/db\_xref="GI:6523480"  
/translation="WTTDKALSRGAYALGGVTPKVTBGTAYVLAGLATCGKWHICP  
CCGAKIRSATVYEQAAGTAMEDIIGINPKTIPAPAAKIDINDTADMAIEAD  
GHOAGLGLMTLTMHYSRHTLAELVTOQORDAMKALQGNAGDKAKDQGVGVF  
RAMEVYGGANGMHPWHYLVFEDPLTPGQDALLEVLYEAMSTLDVGYALPRE  
HGVRLLDSHGSGPRARYLMTYQPKAAWTTAETRTDTAGRGHTTETRIANV  
LTEDAADDRATVLMQETETPAKMRLLVSNGLRKLALVELDTVDGEIAEE  
ROGELAVIADPMHQAARRKRSLLQKAEKGQDQVRALVBSWGVLWGDVLP  
PVDVLEAEQRAEELDRGVGADQORQRAHRAVAKSAVPAKSAAKTAPKKA  
MCTCGGALDRLAGASGAGHIC"  
BASE COUNT 1391 a 3428 c 3248 g 1300 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,44e-19 Length: 9367  
Score: 452.50 Matches: 128  
Percent Similarity: 42.82% Conservative: 54  
Best Local Similarity: 30.12% Mismatches: 158  
Query Match: 22.57% Indels: 85  
Gaps: 14  
US-10-007-527A-2 (1-379) x SNA243257 (1-9367)  
QY 19 ValSerSerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThr 38  
DB 6602 GTGGCGAAGCATCGCGCGATGCGCGTCCGACCTCCGTGAGCGCGCTGACGAGAGC 6661  
QY 39 ThrSerGluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIle 58  
DB 6662 ACGGACAAGCGCGCTGACCGCGCGCGCGGTACCGGCTCGG---GCGGCGTACCGCC 6718  
QY 59 ValAsnGlyProLysGlySerGlyPhe---GlyGlyLeuArgSerCysGlyLysGlyTrp 77  
DB 6719 AAGGTGACCCCGCGGAAACCGCTACCTCGCGGGCTGGCCACCTCGCGGCAAGTGCAC 6778  
QY 78 IleCysProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluLeuSerGlnVal 97  
DB 6779 ATCTGCCCTGCTGCGCGCGAAGATCGCTCGCGCGGTGACGTGCAACTCCAGCGCGCC 6838  
QY 97 ----- 97  
DB 6839 GGGAGCGCTGSGAAGACATCGCGCGCATCAACGCGCGAAGATCCGCCCGCGCGCC 6898  
QY 98 -----ValAla 99  
DB 6899 GCGTGAAGAAAGACATCAACGACCGGAGAGACTGSGCGCGCATCGAGAGAGGCC 6958  
QY 100 HisGlnLeuGlyThrGlySerValAlaMetValThrMetClnMetLysHisThrAlaGly 119  
DB 6959 GATGGGCTCCAGCGCTGTGGCTCGCATGTGACGCTGACCATGCGGACATACCTCCCGG 7018

REFERENCE	AUTHORS	TITLE
JOURNAL	Tauch, A., Pühler, A., Kalinowski, J. and Thierbach, G.	Plasmids in <i>Corynebacterium glutamicum</i> and their molecular classification by comparative genomics
JOURNAL	Tauch, A., Pühler, A., Kalinowski, J. and Thierbach, G.	Submitted (05-NOV-2002) Department of Genetics, University of Bielefeld, Universitätsstrasse 25, Bielefeld D-33615, Germany
FEATURES		<p>Unpublished</p> <p>2 (bases 1 to 4603)</p> <p>Location/Qualifiers</p> <p>1..4603</p> <p>/organism="Corynebacterium glutamicum"</p> <p>/mol_type="genomic DNA"</p> <p>/strain="22220"</p> <p>/db_xref="taxon:1718"</p> <p>/plasmid="pAg3"</p> <p>complement (200..397)</p> <p>/note="orf1"</p> <p>/codon_start=1</p> <p>/evidence=not experimental</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="AA018193.1"</p> <p>/db_xref="GI:27657762"</p> <p>/translation="WSKQKLSPPRAKYLAAVDREYTPVQVAKHKGSTROKAGALMALHKNLIDRTFRGIVALEVDV"</p> <p>complement (613..2061)</p> <p>/gene="repa"</p> <p>complement (613..2061)</p> <p>/gene="repa"</p> <p>/note="replication protein; essential for pAg3 replication"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="Repa"</p> <p>/protein_id="AA018194.1"</p> <p>/db_xref="GI:27657763"</p> <p>/translation="WSILSSKQEAHTSAFNPQATPTHTTGRLGITDTTKHKRYKALS PALFRALMDLQRTMKVYEDKTLACGRVAVGSAVASLVMQAGRAWGIQNSHS WMSPVASVTCSRANVEDAAIOWANGAGLPAHRYGVSTSPDCKGVSPVPLV ERIESTMTLTRENSKOSLTVEWDAIAGCQAVTNTAAWNGARTGDSRGIAHMY BALEVHNGKMGVHVLAVLPHDRVLSVPERDSTLADRVPRMAKAVRIGMRAPSPR GIDVNVAAISDDAKSIGVTCGMLSGIIAETTTGQVTEAKGNRTTPQITGDLCK QYTKRHAIWLEWEGSKGRPQTGWSQGTIDVIGINELSDQIDSLGNDQSEVAM VGGANGAASDTEKRAVILDAVAATDSAGAKARADVILKFGVAHTVAIGLAKTD LDIPLPHTDVFRELBHAWSGRKOTALPVL"</p> <p>2396..2785</p> <p>/note="orf3"</p> <p>/codon_start=1</p> <p>/evidence=not experimental</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="AA018195.1"</p> <p>/db_xref="GI:27657764"</p> <p>/translation="WYRSATCLALHALNVYRCGVNVSEETLLMDTYATLNRICKMAT KVELBOQLDMQALKECOOLAFNAHELEQCRNLKPVYTVAVPQYVRDEMRLLAI QGHTLLOPAAHAFAGITDLSBQDSH"</p> <p>complement (2840..3232)</p> <p>/note="orf4"</p> <p>/codon_start=1</p> <p>/evidence=not experimental</p> <p>/transl_table=11</p> <p>/product="hypothetical protein"</p> <p>/protein_id="AA018196.1"</p> <p>/db_xref="GI:27657765"</p> <p>/translation="MTKPVNRFESANDIEAMIGLVIANNNENYCTDQAKKATAKAIA LNKAADETRRGKDLADDETHIRVPHMLFEQKATHDWQDEVHSRLQSVAHR IEAAGFETRGHMGVQVOLAIFDVE"</p> <p>complement (3344..4030)</p> <p>/note="orfs"</p> <p>/codon_start=1</p> <p>/evidence=not experimental</p> <p>/transl_table=11</p>

```

/product="hypothetical protein"
/protein_id="AA018197.1"
/db_xref="GI:2765776"
/translation="MATPSKPRGLAEQLEKRVSSLELVASNVNADHLSDDKNA
WTRVTAHRTYIAQVEVALETTISMSIDKANDADLPATDTSKVEMELQKTI
NRGIPSVQIGYKLVTSHEPSPTRTALIELEARGLSPEMVGIIKSPVAAATE
MTIQARINSVFCNRLRTMYKALDDKRALFNHVNVTBDSADYDMVEPTVFPFKP
TNAETVTRAS"
/complement(4112, .4519)
/gene="orf6"
/complement(4112, .4519)
/codon="orf6"
/gene="orf6"
/evidence_start=1
/evidence_end=experimental
/translation_table=11
/product="hypothetical protein"
/protein_id="AA018198.1"
/db_xref="GI:27657767"
/translation="WGRNATPAWLVQVSTQIDNISKALNIDMGDHAANASAPAEQ
QSAKLELADKFRKIALDSVLKTDCTTGGERRDQIMTAISDKFLXLTSTGAPQ
EYAGQADAAALADRFMYATKAHVHSPDNDH"
BASE COUNT      1077 a      1276 c      1112 g      1138 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      7.84e-18      Length:      4603
Score:          425.50      Matches:      109
Percent Similarity: 49.07%      Conservative: 75
Best Local Similarity: 29.07%      Mismatches: 138
Query Match:      21.22%      Indels:      53
DB:                1          Gaps:      14

```

US-10-007-527a-2 (1-379) x AY172684 (1-4603)

```

QY      2 ThrSerValSerAlaIuhIsleuSerGlyValAspArgProProValIleuValSerSer 21
Db      1971 ACCACTGACACAACTTAAGCAACCTTAAGGCACTATCACCAGCGCTTTTC----- 1918
QY      22 AspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSerGlu 41
Db      1917 -----CGCGCTAAACTTTGGACCTTACACGCGACACATGTGAAAGTCAAGAGTTAA 1864
QY      42 ThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyVal---ThrIleValAsn 60
Db      1863 ACCCTTGGCCGCGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGG 1804
QY      61 GlyProLysGly---SerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrIleCys 79
Db      1803 CAGGACACAGGGGCGTGCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1744
QY      80 ProCysCysAlaGlyValValAlaGlyAlaHisArgAlaAspGluIleSerGlnValVal--- 98
Db      1743 CCGTGTGACATCGCTGTGTATCTGTTCACGTCGGCTGAATGAGGTTATGTCAGCATTCAG 1684
QY      99 -----AlaHisGlnLeuGlyThrGlySer----- 106
Db      1683 GCGTGGCTAATAGTGTGCTCACTCTGCTCATTAACGGGCGGTGTCCACTCCCT 1624
QY      107 -----ValAlaMet 109
Db      1623 AGTATCGGAAACAGCGCGCGGTGTGTGAATCCGGGTGTGAACGGGAGTATTAGCTTGG 1564
QY      110 ValThrMetThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuThrArgLeu 129
Db      1563 ATGACGTTCACCGCTTAAGGCAATACCGACAGCTCCCTTAAGAGGTGTGGACCGTATT 1504
QY      130 SerAlaAlaTyrLysAlaAlaThrAsnGlyArgArgTyrArg----- 143
Db      1503 GCTGCGCTGTGGCAGCGCTTACTTAACACCGCGCGGTGGCGGTGGTGGTGGTGGTGGTGG 1444
QY      144 ThrGluArgGluMetGlyGlyCysAspGlyTyrValArgAlaValGluIleThrHisGly 163
Db      1443 GCGGATAAAGTGGCTATGGCATTTGGCCATTTGGTACCGGGCTATTGAGGTACCCACGCGC 1384

```

```

QY      164 LysAsnGlyTyrPheIleValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGlu 183
Db      1383 AAGAAATGGCTGGCATGTTTACCTGATCTGTGTCTTCCATGATCGGCGGTGGTGAATGTT 1324
QY      184 AsnIleLeuGluSerPheSerAspAlaMetPheAspArgTyrThrSerLysLeuValSer 203
Db      1323 GATGAAGAAGATTCCTTGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 1264
QY      204 LeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLysIleGlyGly 223
Db      1263 CTGTGATGATCGTGCACCGCTGTGTATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1207
QY      224 GluAlaAspGlnVal-----LeuAlaAlaTyrLeuThrVal---IleAlaSerGlyVal 240
Db      1206 TCGTGTGATGACGCTTAAGCAATGTGTGCTTACCTGTAAAGGCAATGTGTGCTGATATT 1147
QY      241 GlyMetGluValGlySerGlyAsp---GlyLysSerGlyArgHisGlyAsnArgAlaPro 259
Db      1146 GCTGCGGAGACCAACACAGGCGAGGTAACTAAGAAAGCTAAGGAGATTAACCGACACT 1087
QY      260 TyrGluIleAlaValAlaAspAlaValGlyGlyAspProGlnAlaLeuGlu-----LeuTyr 277
Db      1086 TTTCAGATATCTCGGTGAC---CTGGGAAAACAGTACACAAACGCTGATCATGCCCTATG 1030
QY      278 ArgGluPheGluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArg 297
Db      1029 CTGAGTGGGAAAAGGTTCTAAGGCGAGCGGACAGCTGTGTGTCAACAGGCACTAA 970
QY      298 AlaArgAlaGlyLeuGlyValAlaGluLeuThrAspAlaGlnIleValGlnGln-----Glu 315
Db      969 GATGTTCTTGAATTAAAC---GACTGTGATGATGATGATGATGATGATGATGATGATGAT 913
QY      316 GluSerAlaProValMetValAlaIleLeuProAlaArgSerTyr 330
Db      912 AACGATCACTCCGAAGTGTGCTATGTTGGCGCGGAGCGTGG 868

RESULT 13
E17316
LOCUS      E17316      2297 bp      DNA      linear      PART 28-UTL-1999
DEFINITION Bifidobacterium breve gene for replication protein of pNB1.
ACCESSION E17316
VERSION E17316.1 GI:5711999
KEYWORDS JP 199826270-A/1.
SOURCE Bifidobacterium breve
ORGANISM Bifidobacterium breve
Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
Bifidobacteriaceae; Bifidobacterium.
REFERENCE 1 (bases 1 to 2297)
AUTHORS Iino,T. and Morishita,T.
TITLE SHUTTLE VECTOR FOR BIFIDOBACTERIA AND REPLICATION PROTEIN GENE OF
BIFIDOBACTERIA PLASMID
JOURNAL Patent: JP 199826270-A 1 06-OCT-1998;
YAKULT HONSHA CO LTD
OS Bifidobacterium breve
PN JP 199826270-A/1
PD 06-OCT-1998
PT 27-MAR-1997 JP 1997091387
PI IINO TORU, MORISHITA TAKASHI
PC C12N15/09, (C12N15/09, C12R1:19), (C12N15/09, C12R1:01); CC
strandedness: Double;
CC topology: linear;
FH Key Location/Qualifiers
FT source 1..2297
FT /organism='Bifidobacterium breve' FT
FT /strain='ATCC15698'
FT /plasmid='pNB1'
FT -35_signal 387..392
FT -10_signal 410..415
FT RBS 525..530
FT CDS 539..2023
FT /product='replication protein of pNB1'.

```



```

gene
CDS
    /product="partitioning protein"
    /protein_id="AAD34708.1"
    /db_xref="GI:4972587"
    /translation="MILSIASKGSGTGTSSIFLACAAHTADBSLITLLDADPOGS
ATIDAEAEONKDDPLPFMYKYAGESDINLKDNBSGLTIIDTAPGTTDLTOTIKSDI
CITTSPTVIDARBARAKTFEAIQIRAGGSAIILTPANRTQDFRMDIDDLANDANNC
ILFDVIRQSTRYSKPFCHTPKTYDVAADNQEIQKAGS"
    complement(2373..3515)
    /gene="rep"
    /note="replication gene"
    complement(2373..3515)
    /gene="rep"
    /note="rep: similar to replication proteins of
    streptomyces plasmids and other members of the pC194
    family that replicate by a rolling circle mechanism"
    /codon_start=1
    /transl_table=11
    /product="replication protein"
    /protein_id="AAD34709.1"
    /db_xref="GI:4972588"
    /translation="MTNGEKARFTGTMLCGSIWACPTCSAIRHERAEVALAIGNHA
EKLKAADQWQAEHEGQRLPELWSPDSFGNYIFGLTLEHDTMPLAMTLDAILKG
WTKINGSPEORASERAKIRGFVAIETTVGNMHHPIHVMFLDGLDGGSEMOO
OMLDRMKTWYKRAVAKYKKKGDPNVAVADENGLDLPFGSGDAGTAAEYITTKQ
GDKRGVTLAOLAIRAGDILKNGMGSVNPQLDSCGLSDPQREDYMLWOTIRRR
CITSRGCKEDMEVBELEDELAKRABELPELVGVVENVRYKDIRSAPELTADALD
AAEEDMQEVARLLPGVITLDEEQDADAIDAEARPGDYLPTMSYVW"
    complement(3926..3934)
    /note="region conforming to double stranded origin,
    nicking site of rolling circle plasmids"
    complement(3985..4845)
    /gene="f1sk gene homolog"
    complement(3985..4845)
    /gene="f1sk gene homolog"
    /note="hypothetical protein; similar to transfer proteins
    and f1sk homologs"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAD34710.1"
    /db_xref="GI:4972589"
    /translation="MKKPMVNTPEPLPRFLRPFDPATGIGITVGLKANGDNLITV
NTSGVVGGIPGSGKRTAGMVAIVTALVSGCNTHVIDKGGDWMGFSHATFPVG
DLDTVDHTLRLDDEMCKRIASMGQYSANVNVODKPPLEVIIDICQSFPAK
GILGGPKADKABEITAAATEIVQKGSGLFLFAIYOKPDTSLPALENCENRIG
FRVTPPEAARAVLGDMDPSPSPFDIPPARGAILGLATGEDVMCRFAVSEEAR
AVANHHGV"
BASE COUNT      1278 a   1656 c   1616 g   1200 t
ORIGIN

Alignment Scores:
Pred. No.:      1.26e-15      Length:      5750
Score:          392.50      Matches:      106
Percent Similarity: 42.71%      Conservative: 61
Best Local Similarity: 27.11%      Mismatches: 161
Query Match:      19.58%      Indels:      63
DB:              12      Gaps:      12

US-10-007-527A-2 (1-379) x AF085719 (1-5750)

QY      25  G|Y|L|E|A|R|G|H|S|G|U|L|E|U|A|R|G|P|r|o|L|y|L|e|u|G|I|n|I|L|E|T|H|T|R|S|E|R|G|U|T|R|H|P|H|E|A|S|N      44
      |||||
Db      3641 GGTGGCCGCGCTGAAAGCCGAAAGATTCCTGTCGTCATGCGCGGATGCGCAAGCGGTCGGA      3582

QY      45  -----A|A|C|G|L|Y|A|R|P|r|o|L|I|E|S|R|G|Y|V|A|L|S|N      54
      |||||
Db      3581 TTGAGCCGATTAAGCTGCGCGCTGGCGACGCTGCGCCAAACCGGTCGAC-----ACC      3528

QY      55  G|Y|V|A|T|H|T|R|L|E|V|A|L|A|E|N|G|Y|P|r|o|L|y|G|L|Y|S|R|G|Y|P|H|e|G|L|Y|L|E|U|A|R|G|S|R|C|Y|S|G|L|Y      74
      |||||
Db      3527 GGGGTGGGTGTCATGACCAACGCGGAAAGCCCGGTTTACAGGCAACCATGCTGTGCGGC      3468

QY      75  L|Y|E|G|L|Y|T|R|P|L|E|C|Y|S|P|r|o|C|Y|S|A|L|A|G|L|Y|S|V|A|G|L|Y|A|L|H|I|S|R|G|A|L|A|E|R|G|U|L|E      94

```

Db	3467	TCGATCTGGGACATGGCCCACTGGCTCGGCATTCATTCGSCACGAACGGGCCACGACGCTC	3408
Qy	95	SerGlnValValAlaHisGlnLeu-----	102
Db	3407	GCCTCGCCATCGGGAACCATGGACCGAATACTGAGAAAGCCGCCGACCAATGCGAG	3348
Qy	103	-----	glyThrGlySer 106
Db	3347	GCAAGACATGAGGGGACGCTTCGCCCGGAACTCATGTGTCCGACAGCTTGGAAC	3288
Qy	107	ValAlaMetValThrMetThrMetArgHisIleThrAlaGlyGlnArgLeuHisAspLeuTrp	126
Db	3287	TACATCTTCGGACACCTCCACCTCGGCATGATGAGCAATGCCGCTGGCCATGACCTT	3228
Qy	127	ThrGlyLeuSerAlaAlaTrpLysAlaAlaThrAsnGlyArgArgGTPArgThrGluArg	146
Db	3227	GACGCAATCTCGAAGGGCTGGACGAGAATGATTAACGGAACCCCTGGCAACGGGCTCG	3168
Qy	147	GlnMetTrpGlyCysAspGlyTrpValArgAlaValGluIleThrHisGlyLysAsnGly	166
Db	3167	GAACTTGGAATATCAGAGGCTTTGTGTCCGCGCATTAATACACCTACGGGTGAGACGGC	3108
Qy	167	TrpHisValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeu	186
Db	3107	TGGCACTTCACATTCATTCCTGCTATTTCTTCATGGCGATCTGACAGCGGACGAGCT	3048
Qy	187	GluSerPheSerAspAlaMetPheAspArgTrpThrSer-----LysLeuValSerLeu	204
Db	3047	GAGCAATGCAAGCAATGGCTGTGCTGATGCTGGAACCAATGATCAAGCGGCTTGCAAG	2988
Qy	205	GlyPhe-----AlaAlaProLeuArgAsnSerGlyVal	215
Db	2987	GCATACAAAGAAAAAGACGGCAACCCCTACAGCTGCCCGC---AAAGACAAACACGGC	2931
Qy	216	LeuAspValArg---LysIleGlyGlyLysAlaAspGlnValLeuAlaAlaTrpLeuThr	234
Db	2930	ATAATCTTGGCAATTCAGATCGGGCAAGGACGCGCAACCGCTGCGCCGATACATACAC	2871
Qy	235	LysIleAlaSerGlyValGly-----MetGluValGlySerGlyAspGly	249
Db	2870	AAAGATTCAAGCGCAAAAGCGGGCTGACATCTGGCTCAGAAATCGCGCGGAGGATATC	2811
Qy	250	LysSerGlyArgHisGlyAsnArgAlaProTrpGluIleAlaValAspAlaValGlyGly	269
Db	2810	AAGATGTGCTGTAAGGGGTGGGTAAACCCGTTCAAATTGCTGACCTCGGGTGGCTGGG	2751
Qy	270	-----AspProGlnAlaLeuGluLeuTrpArgGluPheGluPheGlySerMetGlyArg	287
Db	2750	CTGTCCCAATTTCCAGCCCGCAAGATCTTGTGCTGAAATACGGCAAGCCACTCTGGCGCGC	2691
Qy	288	ArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGluLeuThr	307
Db	2690	CGTGCATTAACATGTGTGGTGGCTCTCAAGAAAGACATGAGGTC---GAGAACTTGAA	2634
Qy	308	AspAlaGlnIleValGluGlnGluGluSerAlaProValMetValAla---IleIlePro	326
Db	2633	GACACGAGAGCTGGCGGAGAAAGCCGACGAACCTGCCGATCTGTGCTGGCTAATGCTGCGC	2574
Qy	327	AlaArgSerTrpMetCetiLeaTrpThrCysAlaProGlyArgIleGlyGluIleLeuGly	346
Db	2573	AATCGGGTTTACAAAGACATTCGCAAGAGTGGCGCTTGAGACACTGGCCGACGCAATGGAT	2514
Qy	347	LeuValGlnAlaGlyAlaThrTrpGluAsnLeu	357
Db	2513	GCCGCCGAACGC---GAAGACTGGCAGGAAGTC	2484







Db 4399 GGCATATCATTTGATCGGGCTATTCCGGCTTCAGCATTCGAATTCGGTGGCTTGC 4458  
QY ProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAla 99  
Db 4459 CCGGTGTGTCTGCCGATGCTTCTCAGCGGCAATGCTGTGACATGAGTCTGAGC 4518  
QY 100 -----HisGlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThr 117  
Db 4519 AGGTGGCATGCGCAGGG-----GGCCGGGCGATGATGATCAAGCTCACGGTCCGACAGAT 4575  
QY 118 AlaGlyGlnArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrPylsAlaAlaThr 137  
Db 4576 CGCAGCGACATCTCAAAATCGTGTGGGACGCGCTGCTAAGGGCTGTGCAAAAGCCACG 4635  
QY 138 AsnGlyArgArgTyrPyrThrGlnArgGlnMetTyrGlyCysAspGly-----Tyr 154  
Db 4636 AATGTCTGCTCTGGATGTCTTCGGGTCACTGTGTGGGGTGGATGTCGTCTGCCCTTG 4695  
QY 155 ValArgAlaValGluIleThrHisGlyLysAsnGlyTyrHisValHisValHisAlaLeu 174  
Db 4696 CTGGGTTCGTGGAGTGAAGCAGCAGCATCCGGTGGCACTGGCACTGTCATGCGCTG 4755  
QY 175 LeuMetPheSerGlyAspVal-----SerGlnAsnIleLeuGlnSerPheSerAsp 191  
Db 4756 GTCTCTCTGGGCGAGGGGTGCGCAGATCGCAGAGCGACGCGTGGAGCAGATTCCGCCA 4815  
QY 192 AlaMetPheAspArgTyrThrPheSerLysLeuValSerLeuGlyPheAlaAlaProLeuArg 211  
Db 4816 CGCATGTGGGGTCTTGGAATCGTCCGTGTGTGCTCAGGGGTGGAAGTCAGGGGCTGGAA 4875  
QY 212 Asn---SerGlyLeuAspValArgLysIleGlyGluAlaAspGlnValLeuAla 230  
Db 4876 AGCTTGTCAAGGCTCACCTCATGTCAGCCGAGATCCGCGCATGAGCTGGGTGGCTTGCC 4935  
QY 231 AlaTyrLeuThrLysIleAlaSerGlyValGlyMetGlnValGlySerGlyAsp---Gly 249  
Db 4936 GCTTACTTCTTAAGGCCGCT-----TTGAATTGACTCTTGCCG--GGCAGCAAGGC 4988  
QY 250 LysSerGlyArg-----HisGlyAsnArgAlaPro----- 259  
Db 4989 CGCTCAGGGGCAATTCGACGCGTTCGAGATCTCGCGCTCTCACTGGCGACCTCAA 5048  
QY 260 TyrGluIleAlaValAspAlaValGly---GlyAspProGlnAlaLeu-GluLeuTyrPhe 278  
Db 5049 TGGCGAGGTGGGCTGAGACTGTTCGGCAAGAGCGACGCGCTGTGGCGCTGTGGCA 5108  
QY 278 GlnPheGlnPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAl 298  
Db 5109 CGAGTGGAAAGATTCTCGTCCGCTCGTCAGATGGCGCTCCAGGGCCTGTTCGA 5168  
QY 298 aaArgAlaGlyLeuGlyAlaGluLeuThr-----AspAlaGlnIleValGluGlnGlu 316  
Db 5169 CATGCTCGGTCTGATTCGTCTGACGCCCTCGAGATGACGAGATTATCGAGACATAGA 5228  
QY 316 uSerAlaProValMetValAlaIleIleProAlaArgSerTyrMetMetIleArgThrCy 336  
Db 5229 CCTCGATGGGAGAGACTTTCGTGATCGTCAAGAGAAATGGAACGCGCTGCGGAGAC-- 5286  
QY 336 salaproTyrValPheGlyGluIleLeuGlyLeuValGluAlaGlyAla 352  
Db 5287 -----CGTGGCCTGGCTCTCGACGTGCTTGAAGCGCGTTCGT 5322

Search completed: August 22, 2003, 11:37:16  
Job time : 3146 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 22, 2003, 09:08:23 ; Search time 264 Seconds

(without alignments)  
3875.330 Million cell updates/sec

Title: US-10-007-527A-2

Perfect score: 2005

Sequence: 1 MTSVSAHSLGKDRPPVLVS.....HLHRLPAADVPPITSVRK 379.

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xjh  
-O=/cgmt\_1/USPRO.spool/US10007527/runat 20082003 165151 15356/app.query.fasta\_1.519  
-DB=N Geneseq 19Jun03 -OPMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdt  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MOB=LOCAL -OUTPRT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US10007527 @CGN 1.1 0 @runat 20082003 165151 15356 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOUDRY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGBEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_19Jun03: +  
1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: \*  
4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: \*  
8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: \*  
12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: \*  
13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: \*  
18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT: \*  
22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: \*  
25: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2005	100.0	1140	24	ABQ76122
2	2005	100.0	6334	24	Rhodococcus AN12 r
3	2005	100.0	9652	24	Rhodococcus AN12 d
4	2005	100.0	11241	24	plasmid pRRH171 D
5	399.5	19.9	2297	19	plasmid pRRH171 DN
6	337	16.8	9431	13	B. breve essential
7	326	16.3	4447	21	DNA encoding the c
8	326	16.3	4447	21	B. lactofermentum
9	326	16.3	4447	24	B. lactofermentum
10	318	15.9	4447	21	B. lactofermentum
11	318	15.9	4447	21	B. lactofermentum
12	318	15.9	4447	24	B. lactofermentum
13	296	14.8	8500	24	B. lactofermentum
14	282	14.1	4005	24	B. lactofermentum
15	282	13.6	3741	21	B. lactofermentum
16	273	13.6	3741	21	B. lactofermentum
17	273	13.6	3741	21	B. lactofermentum
18	214	10.7	5648	21	B. lactofermentum
19	133	6.6	349980	24	B. lactofermentum
20	127	6.3	3331	10	B. lactofermentum
21	125	6.2	349980	24	B. lactofermentum
22	125	6.2	349980	24	B. lactofermentum
23	125	6.2	349980	24	B. lactofermentum
24	125	6.2	349980	24	B. lactofermentum
25	122.5	6.1	4411529	22	B. lactofermentum
26	122	6.1	2871	23	B. lactofermentum
27	120.5	6.0	4403765	22	B. lactofermentum
28	119	5.9	1910	19	B. lactofermentum
29	118	5.9	15638	25	B. lactofermentum
30	117	5.8	68750	21	B. lactofermentum
31	116	5.8	4466	24	B. lactofermentum
32	116	5.8	4466	24	B. lactofermentum
33	116	5.8	4466	24	B. lactofermentum
34	113	5.6	71989	21	B. lactofermentum
35	113	5.6	349980	24	B. lactofermentum
36	112.5	5.6	1922	19	B. lactofermentum
37	112.5	5.6	4403765	22	B. lactofermentum
38	112.5	5.6	4411529	22	B. lactofermentum
39	112	5.6	4614	22	B. lactofermentum
40	112	5.6	139980	24	B. lactofermentum
41	111.5	5.6	2319	23	B. lactofermentum
42	111	5.5	53789	19	B. lactofermentum
43	111	5.5	349980	24	B. lactofermentum
44	110.5	5.5	2291	9	B. lactofermentum
45	110	5.5	1588	24	B. lactofermentum

#### ALIGNMENTS

##### RESULT 1

ABQ76122 standard; DNA: 1140 BP.

13-JUN-2003 (first entry)

**Rhodococcus AN12 replication protein Rep DNA**

KW Plasmid stability protein; replication protein; ethylene forming enzyme;  
KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
KW polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;  
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase; gene;  
KW shuttle vector; Rep; ds.  
XX Rhodococcus erythropolis.  
OS

XX MO200255709-A2.  
 XX 18-JUL-2002.  
 XX 12-DEC-2001; 2001WO-US47868.  
 XX 12-DEC-2000; 2000US-254868P.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX Bramucci MG, Cheng Q, Kostichka KN, Tomb J;  
 XX WPI; 2002-557827/59.  
 XX P-PSDB; ABB84278.  
 XX New nucleic acid molecule encoding replication protein/plasmid  
 XX stability protein, useful in cloning and expression vectors,  
 XX particularly shuttle vectors for expression of heterologous genes in  
 XX Rhodococcus species -  
 XX Claim 2; Page 63-64; 96pp; English.  
 XX  
 CC This invention describes a novel nucleic acid encoding a replication  
 CC protein or a plasmid stability protein. The product of the invention is  
 CC useful for expression of nucleic acid such as genes encoding enzymes  
 CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
 CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
 CC hydratase, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
 CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
 CC Actinomycetales bacteria. The replication protein or plasmid stability  
 CC protein are useful in cloning and expression vectors and particularly in  
 CC shuttle vectors for the expression of homologous and heterologous genes  
 CC in Rhodococcus sp. This sequence encodes the Rhodococcus AN12 Rep  
 CC protein described in the disclosure of the invention.  
 XX  
 SQ Sequence 1140 BP; 241 A; 287 C; 347 G; 265 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No: 2.24e-156 Length: 1140  
 Score: 2005.00 Matches: 379  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 24 Gaps: 0  
 US-10-007-527A-2 (1-379) x ABQ76122 (1-1140)  
 QY 1 MethionineSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20  
 DB 1 ATGACCAAGCGTAAGTGTGAAACACCTTTCGCGCAAAAGACCGGCTCTCGTGTGCG 60  
 QY 21 SerAspLysArgGlyLeuArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40  
 DB 61 TCCGATTAGCCGCGGATCCGCAAGACTGCGACCCCAACTTCAACAAATACACACGTCA 120  
 QY 41 GluThrPheAsnAlaCysGlyArgProLysSerGlyValAsnGlyValThrIleValAsn 60  
 DB 121 GAACATTATAACGCTGTGGCGGCGGATTTCTGGCGGAGCGGTGACCATTTGTCAAC 180  
 QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTyrIleCysPro 80  
 DB 181 GGTCCGAAAGGTTCTGATTCGAGGCGCTTGTCTTCGCGAAAGGCGTGCATCGCCCC 240  
 QY 81 CysCysAlaGlyLysValGlyAlaHisSerGlyAlaAspGlnIleSerGlnValAlaHis 100  
 DB 241 TGTCTGTGGGAAAAGTCGTCGACATGTGACGAAATTTCTCAAGTTGTGTCTCAT 300  
 QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120  
 DB 301 CAACCTCGGACTGATCTGTGTGCGATGTGTGACATGACCAATGACGCCCATACAGCTGTCA 360  
 QY 121 ArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTyrLysAlaAlaThrAsnGlyArg 140

DB 361 CGGCTCCACGACCTATGAGCTGAGCTTTCGCGACGCTCGCAAGCTCGCAACAGCTCGT 420  
 QY 141 ArgTyrArgThrGluArgGlyMetTyrGlyCysAspGlyTyrValAlaGlyAlaGlnIle 160  
 DB 421 CGTTGGCGTTACGAAACCTGAAATGTACGGCTGCGACGCAATACGTCGCGCTGTGAATC 480  
 QY 161 ThrHisGlyLysAsnGlyTyrPheHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180  
 DB 481 ACTCAGCGAAAAGAAACGCTGCGACAGTCCAGCTTCAACCGCTCTCACTTCAGTGTGAC 540  
 QY 181 ValSerGluSerIleLeuGluSerPheSerAspAlaMetPheAspArgTyrThrSerIys 200  
 DB 541 GTGAGTGAAGAACCTCCGAAATCTCTCGGATGCGATGCTTCGATCGGTGCACTTCCAAA 600  
 QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLys 220  
 DB 601 CTGATATCTCTGGATTTGCTGCGCACCTACGATATTCGGTGTCTCGATGTCGAAGAAG 660  
 QY 221 IleGlyGlyLysAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyVal 240  
 DB 661 ATCGCGGATGAGCTGATCAAGTTCTCGCTCGCATCTGACGAAATTCGATCTGGCGTT 720  
 QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyValArgHisGlyAsnArgAlaProTyr 260  
 DB 721 GGTATGAGGTTGTGTAGTGTGCGACGCAAAAAGTGTGTGACATGGCAACCGTGCACCTGG 780  
 QY 261 GluIleAlaValAspAlaValGlyValAspProGlnAlaLeuGluLeuTyrPheGluPhe 280  
 DB 781 GAATTCCTCTGTGTGAGTGGCGGGGATTCACAAACGTTGGAACTGTGGCGCAATTTT 840  
 QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAlaArgAla 300  
 DB 841 GAGTTTGTTGANTGGACCTCGGCGCAATCGCGTGTCTCGGTGATTCGCGCCGAGCT 900  
 QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGlnGlnGlnSerAlaProVal 320  
 DB 901 GGTCTTGCGGACAGACTMAAGATGCTCAAGATCTTGACGAGAAATCTGCCCGCTC 960  
 QY 321 MetValAlaIleIlePheProAlaArgSerTyrMetMetIleArgThrCysAlaProTyrVal 340  
 DB 961 ATGGTTTCGATCATTTCCGCGCGCATCGTGAATGATGATTCGCACTTGTGCGCTTACGTC 1020  
 QY 341 PheGlyGluIleLeuGlyLeuValGlyAlaGlyValaThrTyrPheGluAsnLeuArgAspHis 360  
 DB 1021 TTGGCGGAGATTCCTCGGACCTGTGCAAGCTGCGCCGACTTGGAAAATCTTCGATCAC 1080  
 QY 361 LeuHisTyrArgLeuProAlaAlaAspValArgProIleIleSerValArgLys 379  
 DB 1081 TTGCATTATCGATTGCGCGCAGCGGATGTGCGGCCCGCATATATATCGTTCCGAAG 1137  
 RESULT 2  
 ID ABQ76124/c  
 AC ABQ76124 standard; DNA; 6334 BP.  
 AC ABQ76124;  
 DT 13-JAN-2003 (first entry)  
 XX  
 DE Rhodococcus AN12 derived plasmid pAN12 DNA.  
 XX  
 XX Plasmid stability protein; replication protein; ethylene forming enzyme;  
 XX carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
 XX polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;  
 XX alcohol dehydrogenase; terpene synthase; cholesterol oxidase;  
 XX shuttle vector; circular; ds.  
 XX  
 XX Rhodococcus erythropolis.  
 OS Synthetic.  
 XX  
 XX WO200255709-A2.  
 XX 18-JUL-2002.

XX 12-DEC-2001; 2001WO-US47868.  
PF  
XX  
XX 12-DEC-2000; 2000US-254868P.  
PR  
XX  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
PA  
XX  
XX Bramucci MG, Cheng Q, Kostichka KN, Tomb J;  
XX WPI; 2002-557827/59.  
XX  
XX  
XX New nucleic acid molecule encoding replication protein/plasmid  
PT stability protein, useful in cloning and expression vectors,  
PT particularly shuttle vectors for expression of heterologous genes in  
PT Rhodococcus species -  
XX  
XX  
XX Claim 19; Page 68-71; 96pp; English.

CC This invention describes a novel nucleic acid encoding a replication  
CC protein or a plasmid stability protein. The product of the invention is  
CC useful for expression of nucleic acid such as genes encoding enzymes  
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
CC Actinomycetales bacteria. The replication protein or plasmid stability  
CC protein are useful in cloning and expression vectors and particularly in  
CC shuttle vectors for the expression of homologous and heterologous genes  
CC in Rhodococcus sp. This sequence represents the Rhodococcus AN12 derived  
CC plasmid PAN12 DNA described in the disclosure of the invention.

XX  
SQ Sequence 6334 BP; 1386 A; 1807 C; 1681 G; 1460 T; 0 other;

#### Alignment Scores:

Pred. No.:	1,76e-155	Length:	6334
Score:	2005.00	Matches:	379
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-007-527a-2 (1-379) x ABQ76124 (1-6334)

QY 1 MetThrseryAlserAlaGluHlsLeuSerGlyLysAspArgProProValLeuValSer 20  
DB 3051 ATGACGACGCTAAAGTGTGTAACACCTTCGCGGAAACCGGCTCCCGTCCTGTCG 2992  
QY 21 SerAspLysArgGlyLysLeuArgHlsGluLeuArgProLysLeuGlnGlnIleThrThrSer 40  
DB 2991 TCCGATTAAGCGCGCATCCGCGACGAACTCGACCCAACTTCACAAATCACCAACGTC 2932  
QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60  
DB 2931 GAAACATTTAACCCCTGTGCGCGCGCATTCGTGCGTGAACGCTGACCATTTGTCAC 2872  
QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTPIIleCysPro 80  
DB 2871 GGTCCGAAAGGTTCTGATTCGAGGCTTCGTTCTCGCGGAAAGGCTGATTCGCCCC 2812  
QY 81 CysCysAlaGlyLysValGlyAlaHlsArgAlaAspGlnIleSerGlnValValAlaHls 100  
DB 2811 TGTCTGTCGGGAAAGTCGCTGCACATCGTCAGACGAAATTTCTCAAGTTGTTGCTCAT 2752  
QY 101 GluLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHlsThrAlaGlyGln 120  
DB 2751 CAACCTGGAGCTGATCTGTGCGATGTAACATGACCAATGCGCATACAGCTGTCAG 2692  
QY 121 ArgLeuHlsAspLeuTrpThrGlyLysSerAlaAlaTrpLysAlaAlaThrAsnGlyArg 140  
DB 2691 CGGCTCACGACCTAAGACTGACCTTCGCGACGCTCGGAAACCTCGACCAACGCTCGT 2632  
QY 141 ArgTrpArgThrGluArgGluMetTrpGlyCysAspGlyTrpValArgAlaValGlnIle 160

DB 2631 CGTTGCGGTACGAAACCTGAATGTACGGCTGCGACGGAATACGTGCGCTGTGAAATC 2572  
QY 161 ThrHlsGlyLysAsnGlyTPIIleValHlsValHlsAlaLeuMetPheSerGlyAsp 180  
DB 2571 ACTACCGGAAAGGCTGCGACGCTCACAGTTCACGCGCTACCTCATGTTCATAGTGTAC 2512  
QY 181 ValSerGluAsnIleLeuGlnSerPheSerAspAlaMetPheAspArgTrpThrSerLys 200  
DB 2511 GTGAGTGAGAACATCCCTGCAATCCTTCGGAATCCGATTCGATTCGATTCGATTCGAAA 2452  
QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysLeuAspValArgLys 220  
DB 2451 CTGCTATCTCTGGGATTTGCTGCGCCACTACGTAATTCGGGTGCTCGATTCGAAAG 2392  
QY 221 IleGlyGlyGluAlaAspGlnValAlaAlaIleThrLeuThrLysIleAlaSerGlyVal 240  
DB 2391 ATCGCGGATGAGTGTATCAAGTTCCTGCGCGATTCGAAATTCGATTCGCGGCTT 2332  
QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyValArgHlsGlyAsnArgAlaProTrp 260  
DB 2331 GGTATGAGGTTGGTGTAGTGGCGACGGAAGAAAGTGTGCAATGCAACCTGCAACCTTGG 2272  
QY 261 GluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeuGlnLeuTrpArgGluPhe 280  
DB 2271 GAAATCGCTGTGTATGACGTGGCGGCGATCCACACGTTGGAACTGTGGCGGAAATTT 2212  
QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuAlaArgAla 300  
DB 2211 GAGTTTGTTGCAAGGACGCTCGGCGCATCGCGTGTGCTCGGTGATTCGCGCGGAGCT 2152  
QY 301 GlyLeuGlyValGluLeuThrAspAlaGlnIleValGluGlnGlnGlnSerAlaProVal 320  
DB 2151 GGTCTTGGGCGCAACATCAACATGCTCAATGCTTACGACGAAAGATTCGCCCGGCTC 2092  
QY 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProLysVal 340  
DB 2091 ATGTTGCGCATTCATTCGCGCGCATCGTGTGATGATGATTCGACTTGTGCGCTTACGTC 2032  
QY 341 PheGlyGluIleLeuGlyLeuValGlyAlaGlyAlaThrTrpGluAsnLeuArgAspHis 360  
DB 2031 TTCGGCGAGATCTCGGACCTGTCGAAAGCTGGCGCATTCGGGAAATTCCTCGATCAC 1972  
QY 361 LeuHlsTrpArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379  
DB 1971 TTGCAATTATGATTCGCCCGCGACGATGTGCGGCCCCCGAATATGCTTCGCAAG 1915  
RESULT 3  
ID ABQ76126 standard; DNA; 9652 BP.  
XX  
AC ABQ76126;  
XX  
XX 13-JAN-2003 (first entry)  
DB  
XX  
XX Plasmid pRHB171 DNA.  
DE  
XX  
XX Plasmid stability protein; replication protein; ethylene forming enzyme;  
KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
KW polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;  
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;  
KW shuttle vector; circular; de.  
XX  
OS Synthetic.  
XX  
XX  
XX WO200255709-A2.  
XX  
PD 18-JUL-2002;  
XX  
XX 12-DEC-2001; 2001WO-US47868.  
PP  
XX  
XX 12-DEC-2000; 2000US-254868P.  
PR  
XX  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX  
PI Bramucci MG, Cheng Q, Kostichka KN, Tomb J;  
XX  
DR WPI; 2002-557827/59.

XX  
PT New nucleic acid molecule encoding replication protein/plasmid  
XX stability protein, useful in cloning and expression vectors,  
XX particularly shuttle vectors for expression of heterologous genes in  
XX Rhodococcus species -  
XX  
XX Claim 27; Page 72; 96pp; English.

XX  
CC This invention describes a novel nucleic acid encoding a replication  
CC protein or a plasmid stability protein. The product of the invention is  
CC useful for expression of nucleic acid such as genes encoding enzymes  
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
CC acid (PHA) synthetases, carotenoid biosynthesis enzymes, nitrile  
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
CC dehydrogenase, terpene synthetases, and cholesterol oxidase in an  
CC Actinomyces bacteria. The replication protein or plasmid stability  
CC protein are useful in cloning and expression vectors and particularly in  
CC shuttle vectors for the expression of homologous and heterologous genes  
CC in Rhodococcus sp. This sequence represents the Plasmid pRHBR171 DNA  
CC described in the disclosure of the invention.

XX  
SQ Sequence 9652 BP; 2097 A; 2729 C; 2559 G; 2267 T; 0 other;

## Alignment Scores:

Pred. No.: 2,91e-155 Length: 9652  
Score: 2005.00 Matches: 379  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 24 Gaps: 0

US-10-007-527A-2 (1-379) x ABQ76126 (1-9652)

QY 1 MetThSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20  
DB 5052 ATGACCGAGCTGAAGTCTGGAACACCTTCCGCGAAAGACCGGCTCCGCTCTGCTCG 4993  
QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnIleThrThrSer 40  
DB 4992 TCCGATTAAGCGCGGATCCGACAGAACTGCGACCAAACTTCAACAATCACCACGTC 4933  
QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60  
DB 4932 GAAACATTTAACGGCTGTGGCGCGGCGGATTTCTGGCGTGAACGGTGTGACCATTC 4873  
QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyThrIleCysPro 80  
DB 4872 GGTCCGAAAGGTTCTGATTCGAGAGGCTTGTCTCTGCGAAAGGCTGATTCGCCCC 4813  
QY 81 CysCysAlaGlyLysValGlyValHisArgHisAspGluIleSerGlnValValHis 100  
DB 4812 TGCCTGTGGGAAAAAGTGGTGCACATGTCGACAGAAATTTCTCAAGTTGCTGCAT 4753  
QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetAlaGHisThrAlaGlyGln 120  
DB 4752 CAACCTGGGAGCTGATCTGTTGGATGGTGAAGTGAACCAAGCCCATACAGCTGCTAG 4693  
QY 121 ArgLeuHisAspLeuThrThrGlyLeuSerAlaAlaThrPheAlaAlaThrAsnGlyArg 140  
DB 4692 CGGCTCCAGCACTGTGACTGAGCTTTCGGCAGGCTTGAAAGCTGCCAACAGGCTGCT 4633  
QY 141 ArgTrpArgThrGluArgGluMetThrGlyCysAspGlyIleValArgAlaValGluIle 160  
DB 4632 CGTTGGCGGTACGGAACGTGAATGTACGGCTGCGACGATACGTCGCGCTGTTGAATC 4573  
QY 161 ThrHisGlyLysAsnGlyThrHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180  
DB 4572 ACTCAGGAAAAAAGCGCTGGCAGCTCCACGTTACCGGCTACATCATGTTACAGTGTGAC 4513  
PT

QY 181 ValSerGluAsnIleLeuGlnSerPheSerAspAlaMetPheAspArgTrpThrSerLys 200  
DB 4512 GTGAGTGAAACATCTCTCGAATCTCTCGAGATGCATGTGATCGGTGCAATTCCTCAAA 4453  
QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLys 220  
DB 4452 CTCGTATCTCTGGAGATTTGCTGGCCACATCGAAATTCGGGTGCTCTCGATGTACGAAAG 4393  
QY 221 IleGlyGlyGluValAspGlnValLeuAlaAlaThrLeuThrValIleAsnSerGlyVal 240  
DB 4392 ATCGCGGAGAACCTGATCAAGTTCTGCTGCGCTGCTGATCTGAAGAAATTCGATCTGGGCTT 4333  
QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrp 260  
DB 4332 GGTATGAGGTTGTGTGTGTGTCGACGGAAGAAAGTGTGACATGCAACCTGCACTCGG 4273  
QY 261 GluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeuGluLeuThrArgGluPhe 280  
DB 4272 GAATCGCTGTGTATGACGTGGCGGGGATCCCAAGCCTTGGAACTGTGGCAGAAATTT 4213  
QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaThrSerArgGlyLeuArgAlaArgAla 300  
DB 4212 GAGTTGTGTGTGATGGGACGTGGGACATGCGTGTGCTCGTGAATTCGTCGCGAGCT 4153  
QY 301 GlyLeuGlyValGluLeuThrAspAlaGlnIleValGluGlnGluSerAlaProVal 320  
DB 4152 GGTCTGGGCGACAACTAACAGATGCTCAGATGCTTGACGACGAAAGATCTGCCCGGCTC 4093  
QY 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTrpVal 340  
DB 4092 ATGGTTGCGATCAATTCGCGCGCATGCTGATGATGATTCGACTTGTGCGCTTACGTC 4033  
QY 341 PheGlyGluIleLeuGlyLeuValGluAlaGlyValThrTrpGluSerLeuArgAspHis 360  
DB 4032 TTCGGCGAGATCTCGGACCTGTGAGAGCTGGGCGCATCTTGGAAGAAATCTTCGATCAC 3973  
QY 361 LeuHisTrpArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379  
DB 3972 TTGATTTATGATTCGCCCGCAGCGAGATGTGCGGCCCGCATATATGCTTCGCAAG 3916  
PT

RESULT 4  
ABQ76125/c  
ID ABQ76125 standard; DNA; 11241 BP.  
XX  
AC ABQ76125;  
XX  
DT 13-JAN-2003 (first entry)  
XX  
DE Plasmid pRHBR17 DNA.  
XX  
KW Plasmid stability protein; replication protein; ethylene forming enzyme;  
KW carotenoid biosynthesis enzyme; isoprenoid, pyruvate decarboxylase;  
KW polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;  
KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;  
KW shuttle vector; circular; ds.  
XX  
OS Synthetic.  
XX  
XX W0200255709-A2.  
XX  
XX 18-JUL-2002.  
XX  
XX 12-DEC-2001; 2001WO-US47868.  
XX  
XX 12-DEC-2000; 2000US-254868P.  
XX  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX  
XX Bramucci MG, Cheng Q, Kostichka KN, Tomb J;  
XX  
XX WPI; 2002-557827/59.  
XX  
XX New nucleic acid molecule encoding replication protein/plasmid

PT stability protein, useful in cloning and expression vectors,  
PT particularly shuttle vectors for expression of heterologous genes in  
Rhodococcus species -  
XX Claim 26, Page 71-72; 96pp; English.  
XX  
CC This invention describes a novel nucleic acid encoding a replication  
CC protein or a plasmid stability protein. The product of the invention is  
CC useful for expression of nucleic acid such as genes encoding enzymes  
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
CC hydratase, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
CC Actinomycetes bacteria. The replication protein or plasmid stability  
CC protein are useful in cloning and expression vectors and particularly in  
CC shuttle vectors for the expression of homologous and heterologous genes  
CC in Rhodococcus sp. This sequence represents the Plasmid pRHB17 DNA  
CC described in the disclosure of the invention.

XX  
SQ Sequence 11241 BP; 2517 A; 3109 C; 2945 G; 2670 T; 0 other;

## Alignment Scores:

Pred. No.:	3,5e-155	Length:	11241
Score:	2005.00	Matches:	379
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-10-007-527A-2 (1-379) x ABQ76125 (1-11241)

QY 1 MetThrSerValSerIaGluHISLeuSerGlyValAspArgProProValLeuValSer 20  
DB 6641 ATACACGAGCTAAGTGTGAACACCTTTCGCGAAAGACCGGCTCCCGCTCGTGC 6582  
QY 21 SerAspLysArgGlyIleArgHISGluLeuArgProLysLeuGlnIleThrTrpSer 40  
DB 6581 TCCGATTAAGCGCGCATCCGCGACGAATCGACCCAACTTAAACAATCACACGTCA 6522  
QY 41 GluThrPheAsnAlaCysGlyArgProLysSerGlyValAsnGlyValThrIleValAsn 60  
DB 6521 GAAACATTTAACCCGTGTGGCCGCGCATTTTCGCGTGAACGCTGACCATTTGCAAC 6462  
QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTrpIleCysPro 80  
DB 6461 GGTCCGAAAGGTTCTGATTCGAGGCTTCGTTCTCGCGAAAGGCTGATCTGCCCC 6402  
QY 81 CysCysAlaGlyLysValGlyAlaHISArgAlaAspGluLysSerGlnValAlaHIS 100  
DB 6401 TGCCTGCGGGAAGAAAGTCGATCGACATCGTCAGACGAATTTCTCAAGTTGCTCAT 6342  
QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHISThrAlaGlyGln 120  
DB 6341 CAACCTCGGAGCTGATCTGTGTGCATGTGATGATGATGATGATGATGATGATGATG 6282  
QY 121 ArgLeuHISAspLeuThrTrpGlyLeuSerAlaAlaTrpIleValAlaAlaThrAsnGlyArg 140  
DB 6281 CGGCTCCACGACCTTAAGATGATGATGATGATGATGATGATGATGATGATGATG 6222  
QY 141 ArgTrpArgThrGluArgGluMetTrpGlyCysAspGlyTrpValArgAlaValGluIle 160  
DB 6221 CGTTGCGTACGGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 6162  
QY 161 ThrHISGlyLysAsnGlyTrpHISValHISValHISAlaLeuLeuMetPheSerGlyAsp 180  
DB 6161 ACTCACGAAAAAAGCGCTGCGACGTCACGCTTCAACGCTTCAATGTTCAATGTTG 6102  
QY 181 ValSerGluAsnIleLeuGlyLysSerPheSerAspAlaMetPheAspArgTrpTrpSerLys 200  
DB 6101 GTGAGTGAAGAACATCTCGATCTTCTCGATGCGATGCGATGCGATGCGATGCGA 6042  
QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValAlaGlyLys 220

DB 6041 CTGATATCTCGGATTTGCTGCGCACACGTAACGTAANTTCGGGTGCTCGATGTACGAAG 5982  
QY 221 ILeGlyGlyGluAlaAspGlnValLeuAlaAlaTrpLeuThrLysIleAlaSerGlyVal 240  
DB 5981 ATCGGCGGTGAAGCTGATCAAGTTCCTGCTGCGTATCTACGAAAAATTCATCTGCGT 5922  
QY 241 GlyMetGluValGlySerGlyLysArgLysSerGlyArgHISGlyValAsnArgAlaProTrp 260  
DB 5921 GGTATGGAAGTGTGATGTGCGACGAAAAAGGTGTGACATGTGCAACCGTGCACCTCG 5862  
QY 261 GluIleAlaValAspAlaValAlaGlyLysAspProGlnAlaLeuGluLeuTrpArgGluPhe 280  
DB 5861 GAAATGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5802  
QY 281 GluPheGlySerMetGlyValArgAlaIleAlaAlaTrpSerArgLysLeuArgAlaArgAla 300  
DB 5801 GATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5742  
QY 301 GlyLeuGlyValGluLeuThrAspAlaGlnIleValGluGlnGluLeuSerAlaProVal 320  
DB 5741 GGTCTTGGGCGAAGATTAACAGATGCTCAGATGCTGAGACAGAAAGATCTGCCGCTC 5682  
QY 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTrpVal 340  
DB 5681 ATGCTTGGCATCATTTCCGCGCGATGCTGATGATGATGATGATGATGATGATGATG 5622  
QY 341 PheGlyGluIleLeuGlyLysValGluAlaGlyAlaIleAlaTrpGluAsnLeuArgAspHIS 360  
DB 5621 TTCGCGAAGATCTCGACTCGTCCGAAAGCTGCGCACTTGCGAAATCTTGATGATC 5562  
QY 361 LeuHISLeuArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379  
DB 5561 TTGCAATATGATGTGCGCGAGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 5505

RESULT 5  
AAVS8945  
ID AAVS8945 standard; DNA; 2297 BP.  
XX  
AC AAVS8945;  
XX  
DT 05-JAN-1999 (first entry)  
XX  
DE B. breve essential region gene.  
XX  
KW Essential region gene; shuttle vector; ds.  
XX  
OS Bifidobacterium breve.  
XX  
FH Key Location/Qualifiers  
FT -35\_signal 387..392  
FT -10\_signal 410..415  
FT RBS 525..530  
FT CDS /\*tag= C  
FT /\*tag= 539..2023  
FT /\*tag= d  
XX  
PN JP10262670-A.  
XX  
PD 06-OCT-1998.  
XX  
PF 27-MAR-1997; 97JP-0091387.  
XX  
PR 27-MAR-1997; 97JP-0091387.  
XX  
PA (HONS ) YAKULT HONSHA KK.  
XX  
DR WPI; 1998-587288/50.  
XX  
DR P-PSDB; AAM73071.  
XX  
PT Shuttle vector for a Bifidobacterium species - contains specific  
PT elements from existing plasmids e.g. pNB1 of ATCC 15698

XX	
PS	Disclosure; Page 14-15; 17pp; Japanese
WV	

CC This sequence represents the essential region gene from plasmid pNBhl  
CC used in the vector of the invention. The vector is a shuttle vector for a  
CC *Bifidobacterium*, and contains: (a) a replicated essential region  
CC originated from a plasmid pNBhl of *Bifidobacterium breve* ATCC 15698;  
CC (b) a replicated essential region originated from a plasmid of *E. coli*;  
CC and (c) an antibiotic-resistant gene originated from a plasmid of *E. coli*  
CC and an antibiotic-resistant gene functioning by a *Bifidobacterium*. The  
CC shuttle vector can be used to transform various *Bifidobacteria* species.

SQ Sequence 2297 BP; 524 A; 659 C; 689 G; 425 T; 0 other;

Alignment Scores:

Pred. No.:	1,78e-23	Length:	2297
Score:	399.50	Matches:	116
Percent Similarity:	42.55%	Conservative:	64
Best Local Similarity:	27.23%	Mismatches:	169
Query Match:	19.93%	Indels:	77
DB:	19	Gaps:	15

US-10-007-527A-2 (1-379) X AAV58945 (1-2297)

QY		4	ValseetlaGlnHisLeuSerGly---	LysAspArgPro-----	GlyIleArgHisGlu	29	15
Db		650	CTAACCCGGGAAGCGGCACGCGGGCGCGTACGGCCGCCCTTGCGAAAAGACGAATAA			709	
QY		16	-----ProValLeuValSerSerAspLysArg-----				
Db		710	ATTAACCCCTCCTCTCCCGACCGCATTTACGGCGCTGGCGGATNAGTGGCGCGGTAA			769	
QY		30	LeuArgProLysLeuGlnGlnIleThrThrseryLysPheAsn-----			44	
Db		770	AGCCGAAAGATTCTGCTCCGTCATCCGGTGGCGCAACCGTCGATTCGAGCCGATTAAG			829	
QY		45	-----AlaCysGlyArgProIleSerGlyValAsnGlyValThrIleVal			59	
Db		830	CTGCCCGCGCTCGCACGCTGGCGGCAACCGGTGAC----	ACCGGCTGTGGGTGCATG		883	
QY		60	AsnGlyProLysGlySerGlyPheGlyGlyLeuArgSerCysGlyLysGlyTrpIleCys			79	
Db		884	ACCAAAGCGGAGAAGACCCCGGTTTACAGGCACCATGCTGTGGGCTCGCATCTGGGATGC			943	
QY		80	ProCysCysAlaGlyLysValGlyValahisArgAlaAspGlnIleSergInValValala			99	
Db		944	CCCACCTGCTGGCAATGATTCGCACAGAACCGCCACAGAGATGCGCTCGCATCGGG			1000	
QY		100	HisGlnLeu-----			102	
Db		1004	AACCATGCCGAGAAACTGAGGAAGACCCGCCCGACCAATGGCAGCGACAATAGAGGG			1066	
QY		103	-----GlyThrGlySerValAlaMetValThr			111	
Db		1064	CAGCGCTCGCCCGCGGAACTCATGCTGTCCGACAGCTTCGGAAACTCACTTTCGGCACC			1122	
QY		112	MethrMetArgHisThrAlaGlyGlnArgLeuhisAspLeuTrpThrGlyLeuSerAla			131	
Db		1124	CTCACCTCGCGGCATGATGGACAAATGCCGCTGGCATGATACCTTGACGCATTCCTCAAG			118	
QY		132	AlaTrpLysAlaIleThrAsnGlyArgArgTrpArgTrpGlnArgGlyMetTrpGlyCys			151	
Db		1184	GCGCTGAGACGAAGATGATTAAACGAAGCCCTTGCGAACCGGCGCTCGGAAGCTTGGAATTC			124	
QY		152	AspGlyTrpValArgAlaValGlnIleThrhisGlyLysAsnGlyTrpHisValHisVal			171	
Db		1244	AGGGGTTTCGTCCGGCGCATTTAAATCACCTTCGGGTGTAAAGGCTGGCACTCCCTCACATT			130	
QY		172	HisAlaLeuLeuMetPheSeryGlyAspValSerGlyAsnIleLeuGlnSerPheSeraSp			191	
Db		1304	CATTTCGCATGTTTCTCGATGGCGATTCGACAGCATGGCGACCGTGAAGCATTCAGCA			136	
QY		192	AlaMetPheAspArgTrpThrser-----LysLeuValSerLeuGlyPhe-----			206	

Db	1364	TGGCTGCTGCATGCTGCTGGAAACCAACATGCTCAACCGCTTGCCAAAGCATATCAAGAAAAA	1423
QY	207	-----Ala1a1aProLeuArgAsnSerGlyIleuAspValArg--	219
Db	1424	GACGGCAACCCCTACCAACGTCGCCG--AAAGCAAGAACACGGCATAGATCTGCATTC	1480
QY	220	Ly1IleGlyIleGlyIleu1a1aAspIleu1a1aLeu1a1aTyrIleuThrIle1a1aSerGly	239
Db	1481	AAATCGGGCAGGAACGCGGAAACCGCTGCGCGCGGAATCAATCAACCAAGATTCAAGCGAC	1540
QY	240	ValGly-----MetGluValGlySerGlyAspGlyLysSerGlyArgHis	254
Db	1541	AAAGCGCGGCTCACTGCTGCTCAGGAATCGCGCGCGCATATCAAGAAATGTCGATATG	1600
QY	255	GlyAsnArgAlaProTyrIleu1a1aVala1aAspAlaValGlyIly-----AspProGln	272
Db	1601	GGGTGCGTAAACCGGTTCGAATGTCTGACTCCGGGTGCTCGGGCTGTCGATTCAG	1660
QY	273	AlaLeuGluLeuThrPheArgIleuPheGlySerMetGlyArgArgAlaIle1a1aTyr	292
Db	1661	CGGAAMATCTCTGGCTCGAATCTGGCAGCGCACTGTGGCGCGCGCTGCTATCAATGG	1720
QY	293	SerArgGlyLeuThrArgAlaArgAlaGlyIleuGlyValaGluLeuThrAspAlaGlnIleVal	312
Db	1721	TGCGTGCTCTCAAGAAACATGGAAGTCT--GAGGAATCGAAAGACGAGAGCTGGCG	1777
QY	313	GluGlnGluIleuSerAlaProValMetValAla---Ile1a1aProAlaArgSerTyrMet	331
Db	1778	GAGAAACCCACCAAGACTGCGCGGTCTGGTGGCTATGTCTGGCCGAATCGGGTTTACAA	1837
QY	332	MetIleArgThrCysAlaProTyrValPheGlyGlyIleuGlyLeuValGluAlaGly	351
Db	1838	GACATTTGCAAGAGTGGCTTGAGACACTGCGCGACATTCGATGCCCGCAAGC--	1894
QY	352	AlaThrTyrGluAsnLeu	357
Db	1895	GAAAGACTGGCAGGAGATC	1912
RESULT 6			
AAQ20907/c	AAQ20907 standard; DNA; 9431 BP.		
ID	AAQ20907		
AC	AAQ20907;		
XX			
XX	22-MAY-1992 (first entry)		
XX			
DE	DNA encoding the chimeric protein VIV2-hCH2-KA.		
XX			
KM	CD4; LTI; Streptococci longisporous; HIV gp120; AIDS; IgG1; T cell;		
KM	immunoglobulin, surface glycoprotein; virus; MHC class II; ss.		
XX			
OS	Homo sapiens.		
XX			
FH	Key		
FT	signal_peptide	648..731	Location/Qualifiers
FT		/*tag= a	
FT	mat_peptide	732..1286	
FT		/*tag= b	
FT		/product= VIV2	
FT	misc_feature	1287..1331	
FT		/*tag= c	
FT		/function= hinge_region	
FT	mat_peptide	1332..1680	
FT		/*tag= d	
FT		/product= CH2	
XX			
XX	WC9200985-A.		
PN			
XX	23-JAN-1992.		
PD			
XX			
PF	01-JUL-1991; 91WO-US04663.		
XX			







PS Example; Page 16-18; 28pp; Japanese.

XX The present invention describes a method for transforming a high  
 CC temperature-resistant coryneform microbe in which the high temperature-  
 CC resistant coryneform microbe is treated with a chemical changing the  
 CC structure of cell wall of the microbe and an electric pulse is applied  
 CC to a solution containing the microbe having a changed structure of cell  
 CC wall and a DNA. Also described are: (1) transforming a high temperature-  
 CC resistant coryneform microbe in which the cell wall of a high  
 CC temperature-resistant coryneform microbe is weakened and electric pulse  
 CC is applied to a solution containing the microbe having weakened cell  
 CC wall and a DNA; (2) a transformant of a high temperature-resistant  
 CC coryneform microbe in which a recombinant DNA is introduced; and  
 CC (3) a transformant of a high temperature-resistant coryneform microbe  
 CC carrying both a plasmid vector derived from the high temperature-  
 CC resistant coryneform microbe and a plasmid vector derived from  
 CC Corynebacterium glutamicum or Brevibacterium lactofermentum. The method  
 CC is used for transforming a high temperature-resistant coryneform microbe.  
 CC The present sequence encodes a Brevibacterium lactofermentum pSRK6  
 CC protein sequence, which is used in an example from the present  
 CC invention.

XX Sequence 4447 BP, 824 A; 997 C; 1378 G; 1248 T; 0 other;

# Alignment Scores:

Pred. No.:	Length:	4447
Score:	Matches:	101
Percent Similarity:	Conservative:	58
Best Local Similarity:	Mismatches:	133
Query Match:	Indels:	42
DB:	Gaps:	14

US-10-007-527a-2 (1-379) x ABL49733 (1-4447)

QY 34 LeuGInGInleThrThrsersGluThrPheAsnAlaCyGlyAlaArgProIleSerGlyVal 53  
 DB 1396 ATGTACAGATCAACCAATGTAAGCTGTGGCGGGGTGCCAAGGTG-GCCAGAGCA 1454  
 QY 54 Aaa-----GlyValThrIleVal-AsnGlyProIleSerGlyPheGlyLeuArg 71  
 DB 1455 AACTGTGGCGGTGCTCGTGGTCTCAACGGT-----GCTTGCAATTGAAGGCTGTGCA 1508  
 QY 71 gSerCyGlyGlyGlyTPrIleCyProCySalGlyValGlyValAlaHisArgAl 91  
 DB 1509 AAACCTCTCACTCTCGTGGGGGTCACTCTGCTGAATTGGAATGCGCAACGCGCG 1568  
 QY 91 aAspGluIleSerGlnValValAlaHisGlnLeuGlyThnGlySerValAlaMetValTh 111  
 DB 1569 CATTAAGCTGCTATTGCTACTAAGATCACTTGGCGGGGTGGCGGCTCATGATGTT 1628  
 QY 111 rMet---ThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTrpThrGlyLeuSe 130  
 DB 1629 TGTGGGCACTGTTCGACACCAACCGGTCAACGATTTGGCGAGGTGAAGCGGTATTAA 1688  
 QY 130 rAlaIaIaTrpIlyAlaAlaIaThrAsnGlyAlaArgTrpArgTrnGluValGlnMetYrGl 150  
 DB 1689 GACTGCGTACTCTCTGATGAGTGAACCAATCTGAGTGAAGAAACGTCACGCGTACCG 1748  
 QY 150 YCyAspGlyTrpValArgAlaValAlaGluIleThrHisGly---LysAsnGlyTrpHisVa 169  
 DB 1749 GGTGAGGACACCTTATGACTATAGGCTCAACACTTTGGCGCGCAAGGTTGGACTT 1808  
 QY 169 lHisValHisAlaLeuLeuMetPheSerGlyAspArgIleSerGluAsnIleLeuGlnSerPh 189  
 DB 1809 GCAACGCAACATGCTGTTGTTCTTGGATGCTGCACCTGTGACAGTGAACGATCAAGCGCT 1868  
 QY 189 eSerAspAlaMetPheAspArgTrpThrSerIleuValSerLeuGlyPheAlaAlaLar 209  
 DB 1869 TGAGGATTCATGATTTTCCCGCTGCTGCTGCTGCTGTTAAGCGCGGATAGACGCGCC 1928  
 QY 209 oLeuArgAsnSerGly---GlyLeuAsp---ValArgIlySllleGlyGlyGluAlaAspGl 227  
 DB 1929 ACTGGGTGAGCAGCGGGGTCAAACTGATGAGGTGCTACTGCGGGTGAAGCGCTGCGAA 1988

QY 227 nValleuAlaAlaIaTrpLeuThrIlySllleAlaSerGlyValGlyMetGluValGlySerGl 247  
 DB 1989 A---ATGGCAACCTTACCTCGCTAAG-----GGCATGTCTCAGAACTGACGTGCTC 2036  
 QY 247 yAspGlyIlySerSerGlyArgHisGlyValAsnArgAlaProTrpGluIleAlaValAspAlaVa 267  
 DB 2037 CGCTACTAAACCGCGCTTAAGGGTGTACACGCGCTTCAGATG---TTGGATATGTT 2093  
 QY 267 lGlyIlyAspProGlnAlaLeuGlu-----LeuTrpArgGl 279  
 DB 2094 GCGCGATCAAGACGACCGCGCGAGATGACACCTGTTTGGTGGCTCGGTGGCTGA 2153  
 QY 279 uPheGluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaAr 299  
 DB 2154 GTATAGGTGTGGTCTTAAAAACCTCGTTG---TCCGTGACGTTGG-----GCTAA 2204  
 QY 229 gAlaIlyleuGlyAlaGluLeuThrAspAlaGlnIle-----ValGluGlnGlu- 315  
 DB 2205 GCGTCTTGGCATATGATTCATATAGACGCTGATGCTGTGAAGTGAAGAAAGAACT 2264  
 QY 316 -----GluSerAlaProValMetValAl 323  
 DB 2265 GTACAGCTCGCGGCTCGAAGACCGGAACGGGTCAATACCCGCTTGTGTTGC 2324  
 QY 323 alleIleProAlaArgSerTrpMetMetIleArgThr 335  
 DB 2325 TTTGGTGAAGCCCGATGATGGAACCTGATTCAGTCT 2361  
 RESULT 9  
 AAD22582  
 ID AAD22582 standard; DNA; 4447 BP.  
 AC AAD22582;  
 XX  
 AC 12-FEB-2002 (first entry)  
 XX  
 DT B. lactofermentum pSRK6 replication control region DNA.  
 DE  
 KW Coryneform bacteria; arginine repressor; L-arginine; liver function;  
 KM amino acid infusion; de.  
 XX  
 OS Brevibacterium lactofermentum.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1318..2598  
 FT /product= "B. lactofermentum pSRK6 replication control  
 FT region Protein"  
 XX  
 PN EP1154020-A2.  
 XX  
 PD 14-NOV-2001.  
 XX  
 XX 24-APR-2001; 2001EP-0109457.  
 PF  
 XX 28-APR-2000; 2000JP-0129167.  
 PR  
 XX (AJIN ) AJINOMOTO KK.  
 PA  
 XX Suga M, Asakura Y, Mori Y, Ito H, Kurahashi O;  
 PI WPI; 2002-012660/02.  
 DR P-PSDB; AAE13530.  
 XX  
 PT Coryneform bacterium with inactivated arginine repressor, useful for  
 PT fermentative production of arginine at high levels -  
 PS Example 1; Page 12-15; 30pp; English.  
 XX  
 CC The present invention relates to Coryneform bacterium in which the  
 CC arginine repressor does not function normally and which produces  
 CC L-arginine. Coryneform bacterium is used to produce L-arginine which is

CC useful in agents for promoting liver function and in amino acid infusions  
CC or in 'comprehensive' amino acid pharmaceutical. Coryneform bacterium  
CC accumulate L-arginine to higher levels than wild-type strains, e.g. 120  
CC compared with 20, mg/dl. The present sequence is  
CC Brevibacterium lactofermentum pSFK6 replication control region DNA.  
CC This DNA is used in the construction of shuttle vector for  
CC Escherichia coli and Coryneform bacteria and temperature sensitive  
CC vector.  
XX

SQ Sequence 4447 BP; 824 A; 997 C; 1378 G; 1248 T; 0 other;

## Alignment Scores:

Pred. No.:	Length:	4447
Score:	4.6e-17	
Percent Similarity:	326.00	Matches: 101
Percent Local Similarity:	47.75%	Conservative: 58
Query Match:	30.33%	Mismatches: 133
DB:	16.26%	Indels: 42
	24	Gaps: 14

US-10-007-527A-2 (1-379) x AAD22582 (1-4447)

```
QY 34 LeuGlnGlnIleThrThrsSerGluThrPheAsnAlaCysGlyAyrProIleSerGlyVal 53
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1396 ATGTACAGATCCACATATGTAAGGCTCTGGCGGGGTGCCATPAGTG-GCGCAGGGACGA 1454
   54 Asn-----GlyValThrIleVal-AsnGlyProIlySerGlyPheGlyIleuAr 71
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1455 AGCTGTGGGTGTCCTGGTGGTCTTAACGGT-----GCTTGCAGATTGAGGGTCTGCA 1508
   71 gSerGlyLyLyGlyTTrIleCyProCyScyAlaGlyLyValGlyAlaHisArgAl 91
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1509 AAACCTCTCACTCGCTGGGGGGGTCACTGGCTGAATTGGAAGTCATGGCGAAGCGCG 1568
   91 aAspGluIleSerGlnValValAlaHisGlnIleuGlyThrGlySerValAlaMetValm 111
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1569 CATTAGCTGCTATTGCTACTATGACATCACTTGGCGGGGTGGCGCCCTCAATGATGTT 1628
   111 rMet---ThrMetArgHisThrAlaGlyIleuArgLeuHisAspLeuThrIleGlyLeuSe 130
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1629 TGTGGGCACTGTTCGACACACCGCTCACAGTCATTTGGCAGGTGGAAGCGGATTTAA 1688
   130 rAlaAlaTrpLySAlaAlaThrAsnGlyAyrArgTrpAyrThrGlyIleuMetTrpG 150
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1689 GACTGGCTACTCTTCATGATGTAAGAAACATCTCAGTGAAGAAAGAACTGCAACGATACG 1748
   150 yCysAspGlyTyTrValAlaValAlaValIleThrHisGly---LySAsnGlyTTrHisVa 169
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1749 GGTGAGACACACCTATATGACTATAGTCAAGACTCTTGGCGCAACGGTTGGACATT 1808
   169 HisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1809 GCAACCGCAACATGCTGTGTTCTTGGATCGTCCACTGCTGACGATGAACCTCAAGCGCTT 1868
   189 eSerAlaAlaMetPheAspArgTrpPheSerLySLeuValSerLeuGlyPheAlaAlaPr 209
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1869 TGAAGATTCATGTTTCCCGCTGGTCTGTGTGTTAAGCGCGGATGGAACCGCGCC 1928
   209 oleuArgAsnSerGly---GlyLeuAsp---ValArgLySileGlyGluAlaAsp 227
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1929 ACTGCTGAGCAGCGGGGTCAAACTTATATAGGTGTTCTACTGGGGTGGAGACGCTCGCA 1988
   227 nValLeuAlaAlaTyTrLeuThrLySileAlaSerGlyValGlyMetGlyValGlySerG 247
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1989 A---ATGGCAACCTACTCGCTAG-----GGCATGTCTCAGAACTGACGTGGCTC 2036
   247 yAspGlyLySerGlyAyrHisGlySnaArgAlaPrTrpGluIleAlaValAlaAspAlaVa 267
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2037 CGCTACTAAACCGCGCTTAAGGGGTCTGACACGCGCTTCAGATG---TTGATATGTGT 2093
   267 lGlyGlyAspProGlnAlaLeuGlyu-----LeuTrpArgG 279
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2094 GGCCGATCAAGGACGCGCGGCGAGATATGACGCTGTTTGGTGGCTCGGTGCGGTGA 2153
```

```
QY 279 uPheGluPheGlySerMetGlyAyrArgAlaIleAlaTrpSerAyrGlyLeuArgAlaAr 299
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2154 GTATAGAGTGTGGTTCTAAAAACCTGCGTTC---TCTGTGTCACGTGG-----GCTAA 2204
   299 GalAaIleuGlyAlaGluLeuThrAspAlaGlnIle-----ValGluGlnGlu-- 315
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2205 GGTGCTTTGGGCATTTATTCATAGACGTGATGCTGTGCTGTAAGAAAGAAAGAACT 2264
   316 -----GluSerAlaProValMetValAl 323
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2265 GTACAGCTCGCGGTCTGGAAGCACCGGAACGGGTCAATATCACCCGGTCTGTGTC 2324
   323 AlIleProAlaAyrSerTrpMetMetIleAyrThr 335
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2325 TTTGTTGAAGCCCGCATGATTTGGAATCATTCATTCAGTCT 2361
   335 TTTGTTGAAGCCCGCATGATTTGGAATCATTCATTCAGTCT 2361
   ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
RESULT 10
AAA90951
ID AAA90951 standard; DNA; 4447 BP.
AC AAA90951;
XX
XX
XX 15-JAN-2001 (first entry)
XX
XX B. lactofermentum pSFK6 coding sequence.
XX
XX B. lactofermentum pSFK6 coding sequence.
XX
XX Temperature sensitive plasmid; TSRCR; protein production;
XX temperature sensitive replication control region; pSFK6; ds.
XX
XX Brevibacterium lactofermentum.
XX
XX
XX Key location/Qualifiers
XX FT CDS 1318..2601
XX FT /*tag= a
XX
XX BP1038966-A1.
XX
XX 27-SEP-2000.
XX
XX 16-MAR-2000; 2000BP-0105326.
XX
XX 16-MAR-1999; 990P-0069896.
XX
XX (AJIN ) AJINOMOTO CO INC.
XX
XX Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX WPI, 2000-573832/54.
XX P-PSDB; AAY97537.
XX
XX Plasmids containing a temperature sensitive replication control regions
XX useful for breeding microorganisms for the production of amino acids by
XX fermentation -
XX
XX Disclosure; Page 19-22; 29pp; English.
XX
XX
XX This sequence encodes the Brevibacterium lactofermentum pSFK6 protein.
XX The invention relates to a plasmid containing a temperature sensitive
XX replication control region (TSRCR) and a marker gene (MG). The TSRCR is
XX derived from plasmid pM330 harboured by Brevibacterium lactofermentum
XX (ATCC 13869) and allows the plasmid to replicate autonomously at a low
XX temperature but does not allow the plasmid to replicate autonomously at
XX an elevated temperature in coryneform bacteria within a temperature
XX range in which the bacteria can grow. The plasmid can be used for
XX modifying a chromosomal gene in a coryneform bacterium, which may be used
XX for the production of useful substances, such as amino acids, by
XX fermentation, to change their genetic traits. Therefore, the plasmid can
XX be used for breeding microorganisms for the production of amino acids by
XX fermentation. The plasmid comprises a TSRCR that allows the plasmid to
XX replicate autonomously at an elevated temperature in coryneform
XX bacteria within a temperature range in which the bacteria can grow.
XX
XX Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;
```



US-10-007-527a-2 (1-379) x ABL49734 (1-4447)

```

QY 34 LeuGInGlnIleThrThrsSerGIuThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
DB 1396 ATGTCAAGATTCACCAATGATGAGGCTCGCGGGGCTGCATAGGTG-GCCGAGGAGACA 1454
QY 54 Aen-----GlyValThrIleVal-AsnGlyProIleSerGlyPheGlyGlyLeuAr 71
DB 1455 AGCTGTGGCGGTGTCTCGGTGCTCAACGGT-----GCTTGCGAGTTGAGGGCTCTGCA 1508
QY 71 gSerCysGlyLysGlyTTPILeCysProCysAlaGlyLysValGlyAlaHisArgAl 91
DB 1509 AAACCTCACTCTCGCTGGGGGTCACTCTGCTGAATGGAAGTCATGGCGCAACCGCG 1568
QY 91 AAspGluIleSerGlnValAlaHisGlnLeuGlyThrGlySerValAlaMetValTh 111
DB 1569 CATTGAGCTGGCTAATGCTACTAAGATACATTGGCGGGGGGTGGCGGCTCATATGTT 1628
QY 111 rMet---ThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTTPThrGlyLeuSe 130
DB 1629 TGTGGGCACTGTTCGACCAACCGCTCAACGATTCATTGGCGAGTTGAAGCGGATTTAA 1688
QY 130 rAlaAlaTTPILysAlaAlaThrAsnGlyArgArgTTPArgThrGluArgGluMetTyrG 150
DB 1689 GACTCGTGAATCTTCTGATGGTGAACATCTTCAGTGAAGAAACGTGACGGTACGG 1748
QY 150 rCysAspGlyTyrValArgAlaValGlnIleThrHisGly--LysAsnGlyTTPHisGly 169
DB 1749 GGTGAGACACCTTATGAGACTATGAGTCAACAGCTTTGGGCGAAGCGTTGGACATT 1808
QY 169 lHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
DB 1809 GCACCGCAACATGCTGTGTTCTTGTGATGTCACAGTCTGACATGAACTCAAGGCAATT 1868
QY 189 eSerAspAlaMetPheAspArgTTPThrseryLysLeuValSerLeuGlyPheAlaAlaPr 209
DB 1869 TGAGGATTCATGATGTTTCCCGCTGCTGCTGCTGCTGTTAAGCGCGGTATGAGCGCGC 1928
QY 209 cLeuArgAsnSerGly---GlyLeuAsp---ValArgLysIleGlyGlyGluAlaAspG 227
DB 1929 ACTGGCTGAGCAGCGGGGTCAACCTGATGAGTGTCTTACCTGGGTGAGAGCGCTGCCGA 1988
QY 227 nValleuAlaAlaTyrLeuThrLysIleAlaSerGlyValGlyMetGluValGlySerG 247
DB 1989 A---ATGGCACTTACTCTGCTAAG-----GGCAGTCTCAGGAACCTACGGGCTC 2036
QY 247 yAspGlyLysSerGlyArgHisGlyAsnArgAlaProTTPGluIleAlaValAspAlaVa 267
DB 2037 CGCTACTAAACCGCGGTCTAAAGGGTCTGACACGCGCTTCAGATG---TTGAAATATGTT 2093
QY 267 lGlyGlyAspProGlnAlaLeuGlu-----LeuTTPArgG 279
DB 2094 GGCCGATCAACGACGACCGCGGAGATATGACGCTGTTTGGTGCTCGGTGGGCTGTA 2153
QY 279 uPheGluPheGlySerMetGlyArgArgAlaIleAlaTTPSerArgGlyLeuArgAlaAr 299
DB 2154 GTATGAGGTGGTGTCTTAAACCTGCTGCG---TCTTGCTACGTTGGG-----GCTAA 2204
QY 299 gAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIle-----ValGluGlnGlu-- 315
DB 2205 GCGTGTCTTGGCATTTGATTACATGACGCTGATGACCTGCTGTAAGTAAGAAAGAACT 2264
QY 316 -----GluSerAlaProValMetValAl 323
DB 2265 GTACAAAGCTCGCGGCTCTGGAAGACCGGACGCGGTGCAATCAACCGCGTTCCTGCT 2324
QY 332 alleleProAlaArgSerTTPMetMetIleArgThr 335
DB 2325 TTTGGTGAAGCCCGAATGATTGGAACATGATTCAGTCT 2361

```

RESULT 12  
AAD22583

```

ID AAD22583 standard; DNA; 4447 BP.
XX AC AAD22583;
XX 12-FEB-2002 (first entry)
DE B. lactofermentum p48K replication control region DNA.
XX Coryneform bacteria; arginine repressor; L-arginine; liver function;
XX amino acid infusion; mutant; ds.
OS Brevibacterium lactofermentum.
XX Key Location/Qualifiers
XX CDS 1318..2598
XX FT /tag= a
XX FT /product= "B. lactofermentum pSFK6 replication control
XX FT region Protein"
XX FT replace (1255, C)
XX FT /tag= b
XX FT replace (1534, C)
XX FT /tag= c
XX FT /note= "This mutation causes an amino acid change
XX FT from proline to serine"
XX FT replace (1866, G)
XX FT /tag= d
XX FT replace (2058, G)
XX FT /tag= e
XX FT replace (2187, C)
XX FT /tag= f
XX FT replace (3193, G)
XX FT /tag= g
XX PN BP1154020-A2.
XX PD 14-NOV-2001.
XX PF 24-APR-2001; 2001EP-0109457.
XX PR 28-APR-2000; 2000JP-0129167.
XX PA (AJIN ) AJINOMOTO KK.
XX PI Suga M, Asakura Y, Mori Y, Ito H, Kurahashi O;
XX WPI; 2002-012660/02.
XX DR P-PsDB; AAB13531.
XX PT Coryneform bacterium with inactivated arginine repressor, useful for
XX fermentative production of arginine at high levels -
XX Example 1; Page 17-20; 30pp; English.
XX PS The present invention relates to Coryneform bacterium in which the
XX CC arginine repressor does not function normally and which produces
XX CC L-arginine. Coryneform bacterium is used to produce L-arginine which is
XX CC useful in agents for promoting liver function and in amino acid infusions
XX CC or in 'comprehensive' amino acid pharmaceutical. Coryneform bacterium
XX CC accumulate L-arginine to higher levels than wild-type strains, e.g. 120
XX CC compared with 20, mg/dl. The present sequence is
XX CC Brevibacterium lactofermentum p48K replication control region DNA.
XX CC This DNA is used in the construction of shuttle vector for
XX CC Escherichia coli and Coryneform bacteria and temperature sensitive
XX CC vector.
XX SQ Sequence 4447 BP; 827 A; 993 C; 1375 G; 1252 T; 0 other;

```

Alignment Scores:

Pred. No.:	2,1e-16	Length:	4447
Score:	318.00	Matches:	100
Percent Similarity:	47.45%	Conservative:	58
Best Local Similarity:	30.03%	Mismatches:	134
Query Match:	15.86%	Indels:	42



```

DB:                24          Gaps:                14
US-10-007-527A-2 (1-379) x AAD22583 (1-4447)
QY      34  LeuGlnGlnIleThrThrsSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
      1396  ATGTACAGATCAACATATGTAAGGCTCTGCGGGGCGCATAGTG- GCGCAGGACGA 1454
QY      54  Asn-----GlyValThrIleVal-AsnGlyProIleSerGlyPheGlyIleuAr 71
      1455  AGCTGTGGCGGTCTCCGTCGTCATACGGT-----GCTTCAGATTGAGGCTCGCA 1508
QY      71  gSerGlyGlyGlyThrIleCysProCysGlyValGlyValAlaHisArgAl 91
      1509  AAACCTCTCACTCGCGGGGGGTCATCTGCTGTAATGGAAGTCATGGCGCAAGCGCG 1568
QY      91  aaSerGluIleSerGlnValValAlaHisGlnLeuGlyThrGlySerValAlaMetValTh 111
      1569  CATTTAGCTGCTGATTCATCTATGAAATCATCTTGGCGGGGCGCGCTCATGATTT 1628
QY      111  rMet---ThrMetArgHisThrAlaGlyIleArgLeuHisAspLeuThrThrGlyLeuSe 130
      1629  TGTGGGCACTGTTCTGACACACCGCTCAGATTCATTGGCGCAGATTGAAGCGGATTAA 1688
QY      130  rAlaAlaTrpValAlaAlaThrAsnGlyArgArgTrpArgThrGluArgGluMetCylrG 150
      1689  GACTGGGTACTCTTCATGATGGTGAACATCTCAGTGAAAGAAACGTGCACGCTACCG 1748
QY      150  yCysAspGlyTyValAlaArgAlaValAlaGluIleThrHisGly---LysAsnGlyTrpHisVa 169
      1749  GGTGGAGCACACCTATAGTACATATAGTACAGACTCTGGGCGCAACGTTGGACATT 1808
QY      169  HisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
      1809  GCACCGCAACATCTGCTGTTGTTCTTGATGTCATCCAGTGTGACAGATGAAGTCAAGCATT 1868
QY      189  eSerAspAlaMetPheAspArgTrpThrSerIleValSerLeuGlyPheAlaAlaTrp 209
      1869  TGAAGATTCATGTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1928
QY      209  oleuArgAsnSerGly---GlyLeuAsp---ValArgLysIleGlyGluAlaAspG 227
      1929  ACTGCTGAGACACGGGCTCAACTATGATCAGGTCCTACTGGGGGAGAGACGCTCGCA 1988
QY      227  nValLeuAlaAlaTrpLeuThrLysIleAlaSerGlyValGlyMetGluValGlySerG 247
      1989  A---ATGCAACCTACTGCTGCTAAG-----GGCATGTCACAGGAAGTCACTGCTGC 2036
QY      247  yAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrpGluIleAlaValAspAlaVa 267
      2037  CGCTACTAAACCGCGCTTAAGAGGCTGTAACGCGCTTCAAGATG---TTGATATATGTT 2093
QY      267  lGlyIlyAspProGlnAlaLeuGlu-----LeuThrArgG 279
      2094  GCGCGCATCAAGAGGACCGCGGCAAGATGACGCTGTTTGTGCTGCTGCTGCTGCTGCTG 2153
QY      279  uPheGluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaAr 299
      2154  GATGAGGTTGGTTCTTAAAAAAGCTGCTG---TCTGTCTACGTTGGG-----GCTAA 2204
QY      299  gAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIle-----ValGluGlnGlu-- 315
      2205  GCGTGTCTTGGGCTATGATTAACATAGACGCTGATGTACGTCGGAATGGAAGAAAGAACT 2264
QY      316  -----GluSerAlaProValMetValAl 323
      2265  GTRCAAGCTCGCGGCTGTGAGACACCGGAACGCGTGGATCAACCGCGCTGCTGTGCTG 2324
QY      323  aIleIleProAlaArgSerTrpMetIleArgThr 335
      2325  TTTGTGAAGCCCGATGATTGGAACCTGATTCACTCT 2361

```

```

ABA93871/C
ID  ABA93871 standard; DNA; 8500 BP.
XX  AC
XX  ABA93871;
XX  DT  02-MAY-2002 (first entry)
XX  DE  E. coli/coryneform bacteria shuttle vector DNA sequence SEQ ID NO:7.
XX  ADH gene; PDC gene; tac promoter; bacteria; shuttle vector; ethanol;
XX  alcohol dehydrogenase; pyruvate decarboxylase; ds.
XX  OS  Escherichia coli.
XX  OS  Coryneform bacterium.
XX  OS  Synthetic.
XX  PN  WO200196573-A1.
XX  PD  20-DEC-2001.
XX  PF  12-JUN-2001; 2001WO-JP04935.
XX  PR  16-JUN-2000; 2000JP-0181625.
XX  PA  (REIN-) RES INST INNOVATIVE TECHNOLOGY EARTH.
XX  PI  Yukawa H;
XX  DR  WPI; 2002-106471/14.
XX  PT  Efficient, high-yielding production of ethanol by recombinant
XX  PT  coryneform bacterium carrying genes expressing pyruvate decarboxylase
XX  PT  and optionally alcohol dehydrogenase activity without substantial
XX  PT  proliferation -
XX  PS  Example 2; Page 24-29; 35pp; Japanese.
XX  CC  The present invention describes a method for producing ethanol using a
XX  CC  coryneform bacterium which is transformed by DNA expressing a pyruvate
XX  CC  decarboxylase activity and if required a gene expressing alcohol
XX  CC  dehydrogenase activity under a regulatory sequence to allow expression
XX  CC  under ethanol production conditions. The bacterium does not proliferate
XX  CC  substantially. Also described are: (1) an expression vector for
XX  CC  transforming the coryneform bacterium which is integrated with a gene
XX  CC  expressing a pyruvate decarboxylase activity and if required a gene
XX  CC  expressing alcohol dehydrogenase activity under a regulatory sequence
XX  CC  to allow the expression for use in the ethanol production; and (2) a
XX  CC  coryneform bacterium transformed with the vector of (1). The method is
XX  CC  used for producing ethanol. The recombinant bacterium can produce
XX  CC  ethanol efficiently with high productivity. The present sequence
XX  CC  represents an E. coli/coryneform bacteria shuttle vector having a PDC
XX  CC  gene linking with a tac promoter and an ADH gene linking with a tac
XX  CC  promoter which is used in an example from the present invention.
XX  SQ  Sequence 8500 BP; 2125 A; 2289 C; 2079 G; 2007 T; 0 other;

Alignment Scores:
Pred. No.:      3e-14      Length:      8500
Score:          296.00     Matches:      105
Percent Similarity: 48.06% Conservative: 56
Best Local Similarity: 31.34% Mismatches: 131
Query Match:      14.76% Indels:      45
DB:              24      Gaps:      14

US-10-007-527A-2 (1-379) x ABA93871 (1-8500)
QY      34  LeuGlnGlnIleThrThrsSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
      7112  ATGTACAGATCAACATATGTAAGGCTCTGCGGGGCGCATAGTG- GCGCAGGACGA 7054
QY      54  Asn-----GlyValThrIleVal-AsnGlyProIleSerGlyPheGlyIleuAr 71
      7053  AGCTGTGGCGGTCTCCGTCGTCATACGGT-----GCTTCAGATTGAGGCTCGCA 7000

```



```

QY 71 gSerCysGlyLysGlyTyrPheCysProCysCysAlaGlyLysValGlyAlaHisArgAl 91
   ::::
Db 6999 AACTCTCACTCTCGCTGGGAGGATTCACCTGCTGATGATGAGATGAGGAGGAGG 6940
QY 91 aAspGluIleSerGlnValAlaHisGlnLeuGlyThrGlySerValAlaMetValThr 111
   ::::
Db 6939 CATTGAGCTGCTATTGCTACTAAGATACATTCGGGCGGCGGCGGCGCTCATGATGTT 6880
QY 111 rMet---ThrMetArgHis-ThrAlaGlyLysArgLeuHisAspLeuTyrThr-GlyLeu 129
   ::::
Db 6879 TGTGGGCACTGTTTCGACACACACCGCTCACAGTCATTTGGCGAGTTTGAAGAGCTGGAT 6820
QY 130 SerAlaIleTyrLysAlaAlaThrAsnGlyArgArgTyrThrGlyLysGlyMetCys 149
   ::::
Db 6819 AAGACTGCTACTCTTCATGATGTAAGAAACATCTCACTGGAAGAAAGACGTCACGCTAC 6760
QY 150 GlyCysAspGlyTyrValArgAlaValGlyIleThrHisGly---LysAsnGlyTyrHis 168
   ::::
Db 6759 GGGGTGGAGCACACCTATAGTACATAGAGGTACACAGACTCTTGGCGACGCTGGGAC 6700
QY 169 ValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGlyLysAsnIleLeuGlySer 188
   ::::
Db 6699 TTGCACCG-CACATGCTGTTGTTCTTGATCGTCCACTGCTGACATGAATCAACTCAAGCG 6641
QY 189 PheSerAspAlaMetPheAspArgTyrThrSerLysLeuValSerLeuGlyPheAlaAla 208
   ::::
Db 6640 TTGAGGATTCATGTTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6581
QY 209 ProlLeuArgAsnSerGly---GlyLeuAsp---ValArgLysIleGlyLysGlyAlaAsp 226
   ::::
Db 6580 CCACTGCGTGAAGACCGGGGTCAAACTTGATCAAGTGTCTCACTGGGTGGAACGCTGGCG 6521
QY 227 GlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyValGlyMetGlyValGlySer 246
   ::::
Db 6520 AAA---ATGCAACCTACCTCGCTAAG-----GGATGCTCTAGAACTGACTGGCG 6473
QY 247 GlyAspGlyLysSerGlyArgHisGlyLysArgAlaProTyrGluIleAlaValAlaAspAla 266
   ::::
Db 6472 TCCGCTACTAAACCGCGCTAAGGGGTGCTGACACCGCTTCAGATG---TTGGATAG 6416
QY 267 ValGlyGlyAspProGlnAlaLeuGly-----LeuTyrArg 278
   ::::
Db 6415 TTGGCCGATCAAGACGACCGCGGAGATATGAGACGTTTGTGGCTCGGTGGCGCT 6356
QY 279 GluPheGluPheGlySerMetGlyArgAlaGlyAlaIleAlaTyrSerArgGlyLeuArgAla 298
   ::::
Db 6355 GAGTATGAGGTGTGTTCTTAAACCTGCGTTG---TCTGTGTCAGTGGG-----GCT 6305
QY 299 ArgAlaGlyLeuGlyValaGluLeuThrAspAlaGlnIle-----ValGluGlnGlu 315
   ::::
Db 6304 AAGCGTCTTTGGGCAATGATTCATATAGACCGCTGATGTAAGTGTGAATGGAAGAAGA 6245
QY 316 -----GluSerAlaProValMetVal 322
   ::::
Db 6244 CTGTACAGCTCGCCGCTGTGAAGACACCGGAACGGGTGGAATCAACCGCGCTTGTGCTTT 6185
QY 323 AlaIleIleProAlaArgSerTyrMetMetIleArgThr 335
   ::::
Db 6184 GCTTGTGTAAGCCCGATGATTTGAAATGATTCAGTCT 6146

```

## RESULT 14

AA517124/c  
ID AA517124 standard; DNA; 2401 BP.

AA517124;

12-MAR-2002 (first entry)

Replicon sequence of Ketogulonigenium endogenous plasmid PADMK612.

Ketogulonigenium; 2-keto-L-gulonic acid; L-sorbose; sorbitol;

KW plasmid replicon PADMK612; replicon; ds.

```

XX 06 Ketogulonigenium robustum strain PADMK61.
XX 07
XX 08
XX 09
XX 10
XX 11
XX 12
XX 13
XX 14
XX 15
XX 16
XX 17
XX 18
XX 19
XX 20
XX 21
XX 22
XX 23
XX 24
XX 25
XX 26
XX 27
XX 28
XX 29
XX 30
XX 31
XX 32
XX 33
XX 34
XX 35
XX 36
XX 37
XX 38
XX 39
XX 40
XX 41
XX 42
XX 43
XX 44
XX 45
XX 46
XX 47
XX 48
XX 49
XX 50
XX 51
XX 52
XX 53
XX 54
XX 55
XX 56
XX 57
XX 58
XX 59
XX 60
XX 61
XX 62
XX 63
XX 64
XX 65
XX 66
XX 67
XX 68
XX 69
XX 70
XX 71
XX 72
XX 73
XX 74
XX 75
XX 76
XX 77
XX 78
XX 79
XX 80
XX 81
XX 82
XX 83
XX 84
XX 85
XX 86
XX 87
XX 88
XX 89
XX 90
XX 91
XX 92
XX 93
XX 94
XX 95
XX 96
XX 97
XX 98
XX 99
XX 100
XX 101
XX 102
XX 103
XX 104
XX 105
XX 106
XX 107
XX 108
XX 109
XX 110
XX 111
XX 112
XX 113
XX 114
XX 115
XX 116
XX 117
XX 118
XX 119
XX 120
XX 121
XX 122
XX 123
XX 124
XX 125
XX 126
XX 127
XX 128
XX 129
XX 130
XX 131
XX 132
XX 133
XX 134
XX 135
XX 136
XX 137
XX 138
XX 139
XX 140
XX 141
XX 142
XX 143
XX 144
XX 145
XX 146
XX 147
XX 148
XX 149
XX 150
XX 151
XX 152
XX 153
XX 154
XX 155
XX 156
XX 157
XX 158
XX 159
XX 160
XX 161
XX 162
XX 163
XX 164
XX 165
XX 166
XX 167
XX 168
XX 169
XX 170
XX 171
XX 172
XX 173
XX 174
XX 175
XX 176
XX 177
XX 178
XX 179
XX 180
XX 181
XX 182
XX 183
XX 184
XX 185
XX 186
XX 187
XX 188
XX 189
XX 190
XX 191
XX 192
XX 193
XX 194
XX 195
XX 196
XX 197
XX 198
XX 199
XX 200
XX 201
XX 202
XX 203
XX 204
XX 205
XX 206
XX 207
XX 208
XX 209
XX 210
XX 211
XX 212
XX 213
XX 214
XX 215
XX 216
XX 217
XX 218
XX 219
XX 220
XX 221
XX 222
XX 223
XX 224
XX 225
XX 226
XX 227
XX 228
XX 229
XX 230
XX 231
XX 232
XX 233
XX 234
XX 235
XX 236
XX 237
XX 238
XX 239
XX 240
XX 241
XX 242
XX 243
XX 244
XX 245
XX 246
XX 247
XX 248
XX 249
XX 250
XX 251
XX 252
XX 253
XX 254
XX 255
XX 256
XX 257
XX 258
XX 259
XX 260
XX 261
XX 262
XX 263
XX 264
XX 265
XX 266
XX 267
XX 268
XX 269
XX 270
XX 271
XX 272
XX 273
XX 274
XX 275
XX 276
XX 277
XX 278
XX 279
XX 280
XX 281
XX 282
XX 283
XX 284
XX 285
XX 286
XX 287
XX 288
XX 289
XX 290
XX 291
XX 292
XX 293
XX 294
XX 295
XX 296
XX 297
XX 298
XX 299
XX 300
XX 301
XX 302
XX 303
XX 304
XX 305
XX 306
XX 307
XX 308
XX 309
XX 310
XX 311
XX 312
XX 313
XX 314
XX 315
XX 316
XX 317
XX 318
XX 319
XX 320
XX 321
XX 322
XX 323
XX 324
XX 325
XX 326
XX 327
XX 328
XX 329
XX 330
XX 331
XX 332
XX 333
XX 334
XX 335
XX 336
XX 337
XX 338
XX 339
XX 340
XX 341
XX 342
XX 343
XX 344
XX 345
XX 346
XX 347
XX 348
XX 349
XX 350
XX 351
XX 352
XX 353
XX 354
XX 355
XX 356
XX 357
XX 358
XX 359
XX 360
XX 361
XX 362
XX 363
XX 364
XX 365
XX 366
XX 367
XX 368
XX 369
XX 370
XX 371
XX 372
XX 373
XX 374
XX 375
XX 376
XX 377
XX 378
XX 379
XX 380
XX 381
XX 382
XX 383
XX 384
XX 385
XX 386
XX 387
XX 388
XX 389
XX 390
XX 391
XX 392
XX 393
XX 394
XX 395
XX 396
XX 397
XX 398
XX 399
XX 400
XX 401
XX 402
XX 403
XX 404
XX 405
XX 406
XX 407
XX 408
XX 409
XX 410
XX 411
XX 412
XX 413
XX 414
XX 415
XX 416
XX 417
XX 418
XX 419
XX 420
XX 421
XX 422
XX 423
XX 424
XX 425
XX 426
XX 427
XX 428
XX 429
XX 430
XX 431
XX 432
XX 433
XX 434
XX 435
XX 436
XX 437
XX 438
XX 439
XX 440
XX 441
XX 442
XX 443
XX 444
XX 445
XX 446
XX 447
XX 448
XX 449
XX 450
XX 451
XX 452
XX 453
XX 454
XX 455
XX 456
XX 457
XX 458
XX 459
XX 460
XX 461
XX 462
XX 463
XX 464
XX 465
XX 466
XX 467
XX 468
XX 469
XX 470
XX 471
XX 472
XX 473
XX 474
XX 475
XX 476
XX 477
XX 478
XX 479
XX 480
XX 481
XX 482
XX 483
XX 484
XX 485
XX 486
XX 487
XX 488
XX 489
XX 490
XX 491
XX 492
XX 493
XX 494
XX 495
XX 496
XX 497
XX 498
XX 499
XX 500
XX 501
XX 502
XX 503
XX 504
XX 505
XX 506
XX 507
XX 508
XX 509
XX 510
XX 511
XX 512
XX 513
XX 514
XX 515
XX 516
XX 517
XX 518
XX 519
XX 520
XX 521
XX 522
XX 523
XX 524
XX 525
XX 526
XX 527
XX 528
XX 529
XX 530
XX 531
XX 532
XX 533
XX 534
XX 535
XX 536
XX 537
XX 538
XX 539
XX 540
XX 541
XX 542
XX 543
XX 544
XX 545
XX 546
XX 547
XX 548
XX 549
XX 550
XX 551
XX 552
XX 553
XX 554
XX 555
XX 556
XX 557
XX 558
XX 559
XX 560
XX 561
XX 562
XX 563
XX 564
XX 565
XX 566
XX 567
XX 568
XX 569
XX 570
XX 571
XX 572
XX 573
XX 574
XX 575
XX 576
XX 577
XX 578
XX 579
XX 580
XX 581
XX 582
XX 583
XX 584
XX 585
XX 586
XX 587
XX 588
XX 589
XX 590
XX 591
XX 592
XX 593
XX 594
XX 595
XX 596
XX 597
XX 598
XX 599
XX 600
XX 601
XX 602
XX 603
XX 604
XX 605
XX 606
XX 607
XX 608
XX 609
XX 610
XX 611
XX 612
XX 613
XX 614
XX 615
XX 616
XX 617
XX 618
XX 619
XX 620
XX 621
XX 622
XX 623
XX 624
XX 625
XX 626
XX 627
XX 628
XX 629
XX 630
XX 631
XX 632
XX 633
XX 634
XX 635
XX 636
XX 637
XX 638
XX 639
XX 640
XX 641
XX 642
XX 643
XX 644
XX 645
XX 646
XX 647
XX 648
XX 649
XX 650
XX 651
XX 652
XX 653
XX 654
XX 655
XX 656
XX 657
XX 658
XX 659
XX 660
XX 661
XX 662
XX 663
XX 664
XX 665
XX 666
XX 667
XX 668
XX 669
XX 670
XX 671
XX 672
XX 673
XX 674
XX 675
XX 676
XX 677
XX 678
XX 679
XX 680
XX 681
XX 682
XX 683
XX 684
XX 685
XX 686
XX 687
XX 688
XX 689
XX 690
XX 691
XX 692
XX 693
XX 694
XX 695
XX 696
XX 697
XX 698
XX 699
XX 700
XX 701
XX 702
XX 703
XX 704
XX 705
XX 706
XX 707
XX 708
XX 709
XX 710
XX 711
XX 712
XX 713
XX 714
XX 715
XX 716
XX 717
XX 718
XX 719
XX 720
XX 721
XX 722
XX 723
XX 724
XX 725
XX 726
XX 727
XX 728
XX 729
XX 730
XX 731
XX 732
XX 733
XX 734
XX 735
XX 736
XX 737
XX 738
XX 739
XX 740
XX 741
XX 742
XX 743
XX 744
XX 745
XX 746
XX 747
XX 748
XX 749
XX 750
XX 751
XX 752
XX 753
XX 754
XX 755
XX 756
XX 757
XX 758
XX 759
XX 760
XX 761
XX 762
XX 763
XX 764
XX 765
XX 766
XX 767
XX 768
XX 769
XX 770
XX 771
XX 772
XX 773
XX 774
XX 775
XX 776
XX 777
XX 778
XX 779
XX 780
XX 781
XX 782
XX 783
XX 784
XX 785
XX 786
XX 787
XX 788
XX 789
XX 790
XX 791
XX 792
XX 793
XX 794
XX 795
XX 796
XX 797
XX 798
XX 799
XX 800
XX 801
XX 802
XX 803
XX 804
XX 805
XX 806
XX 807
XX 808
XX 809
XX 810
XX 811
XX 812
XX 813
XX 814
XX 815
XX 816
XX 817
XX 818
XX 819
XX 820
XX 821
XX 822
XX 823
XX 824
XX 825
XX 826
XX 827
XX 828
XX 829
XX 830
XX 831
XX 832
XX 833
XX 834
XX 835
XX 836
XX 837
XX 838
XX 839
XX 840
XX 841
XX 842
XX 843
XX 844
XX 845
XX 846
XX 847
XX 848
XX 849
XX 850
XX 851
XX 852
XX 853
XX 854
XX 855
XX 856
XX 857
XX 858
XX 859
XX 860
XX 861
XX 862
XX 863
XX 864
XX 865
XX 866
XX 867
XX 868
XX 869
XX 870
XX 871
XX 872
XX 873
XX 874
XX 875
XX 876
XX 877
XX 878
XX 879
XX 880
XX 881
XX 882
XX 883
XX 884
XX 885
XX 886
XX 887
XX 888
XX 889
XX 890
XX 891
XX 892
XX 893
XX 894
XX 895
XX 896
XX 897
XX 898
XX 899
XX 900
XX 901
XX 902
XX 903
XX 904
XX 905
XX 906
XX 907
XX 908
XX 909
XX 910
XX 911
XX 912
XX 913
XX 914
XX 915
XX 916
XX 917
XX 918
XX 919
XX 920
XX 921
XX 922
XX 923
XX 924
XX 925
XX 926
XX 927
XX 928
XX 929
XX 930
XX 931
XX 932
XX 933
XX 934
XX 935
XX 936
XX 937
XX 938
XX 939
XX 940
XX 941
XX 942
XX 943
XX 944
XX 945
XX 946
XX 947
XX 948
XX 949
XX 950
XX 951
XX 952
XX 953
XX 954
XX 955
XX 956
XX 957
XX 958
XX 959
XX 960
XX 961
XX 962
XX 963
XX 964
XX 965
XX 966
XX 967
XX 968
XX 969
XX 970
XX 971
XX 972
XX 973
XX 974
XX 975
XX 976
XX 977
XX 978
XX 979
XX 980
XX 981
XX 982
XX 983
XX 984
XX 985
XX 986
XX 987
XX 988
XX 989
XX 990
XX 991
XX 992
XX 993
XX 994
XX 995
XX 996
XX 997
XX 998
XX 999
XX 1000

```

```
QY 195 AsparGTPThrSerLysLeuValSerLeuGlyPheAlaProLeuArgAsnSergly 214
Db 1175 -----TGGCGTGCATGTCGAAAGGAGTATGCGATGTGGGCGC-----AACGATGCG 1131
QY 215 GlyLeuAspValArgLysIleGlyGlyGluAlaAspGlnValLeuAlaAlaTyrLeuThr 234
Db 1130 GCGTTTGACGTTCGGGGGCTGCCAATGCGGGCGAC-----TAGGTAGCG 1086
QY 235 LysIleAlaSerGlyValGlyMetGluValGlySerglyAspGlyLysSerglyArgHis 254
Db 1085 AAG-----TGGGGCGCTGCCGCAAGAGTTCGCTTCAAGCTCGAAGAGCGGAAACCG 1032
QY 255 GlyAsnArgAlaProTyrGlnIleAlaValAspAlaValGlyLysAspProGlnAlaLeu 274
Db 1031 AAGAGCGACGCCAAGGACGACTACTT-----CAGCGCGCGGATGAC----- 990
QY 275 GluLeuTTPArgGlnPheGluPheGlySermMetGly-----ArgArgAlaIleAlaTyr 292
Db 989 GGGCTTTGGCTCGAATATTTCATGCAACGACGCGCAAGCGTACGCGCAATTGGTGG 930
QY 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGlyLeuThrAsp----- 308
Db 929 AGCCAAAGCTTGAAGAAGATGCGGTTTG---GTGAGCTGATGATGACATGAAGCTATG 873
QY 309 AlaGlnIleValGluGlnGluGlnSerAlaProValMetValAlaIleIleProAlaArg 328
Db 872 GGTGAAGTTGATGACGACGACGAAAGCGGCTCGAATTTGCGGAGTGGGACATGAA 813
QY 329 SerTyrMetMetIleArg 334
Db 812 GGTGAAGCAAGATGCGC 795

RESULT 15
AAS17120/C
ID AAS17120 standard; DNA; 4005 BP.
AC AAS17120;
XX
DT 12-MAR-2002 (first entry)
XX
DE Ketogulonigenium endogenous plasmid pADMX6L2 DNA sequence.
XX
KW Ketogulonigenium; 2-keto-L-gulonic acid; L-sorbose; sorbitol;
KW plasmid pADMX6L2; replication protein; cyclic; circular; ds.
XX
OS Ketogulonigenium robustum strain ADMX6L.
XX
Key Location/Qualifiers
FT CDS complement (3960..2560)
FT FT /*tag= a
FT FT /product= "pADMX6L2 replication protein"
FT FT /note= "pADMX6L2 rep ORF"
FT FT 455..2060
FT FT /*tag= b
FT FT /note= "pADMX6L2 replicon, corresponds to AAS17124"
XX
PD WO200177348-A2.
XX
PD 18-OCT-2001.
XX
PF 05-APR-2001; 2001WO-US11097.
XX
PR 05-APR-2000; 2000US-194627P.
XX
PA (ARCH ) ARCHER-DANIELS MIDLAND CO.
PA (DELI/) D'ELIA J.
PA (STOD/) STODDARD S F.
XX
PI D'Elia J, Stoddard SF;
XX
DR WPI, 2002-041295/05.
DR P-PSDB; AAU1039.
```

```
XX
PR New bacterium of Ketogulonigenium genus, useful for producing
PR 2-keto-L-gulonic acid from sorbose or sorbitol, comprises transgene
PR containing DNA sequence from endogenous Ketogulonigenium plasmid -
XX
PS Claim 10; Fig 2; 116pp; English.
XX
CC The present invention relates to a new bacterium of genus
CC Ketogulonigenium. Ketogulonigenium may further comprise a
CC transgene, comprising a DNA sequence from an endogenous
CC Ketogulonigenium plasmid. Methods for transforming
CC Ketogulonigenium are also described. The invention is useful for
CC producing 2-keto-L-gulonic acid (2-KG) from L-sorbose or sorbitol.
CC The present sequence represents Ketogulonigenium endogenous plasmid
CC pADMX6L2 DNA.
CC Note: AAU1039 is encoded by the reverse complement of bases
CC 3960-2560 of this nucleic acid sequence.
XX
SQ Sequence 4005 BP, 866 A; 1174 C; 950 G; 1015 T; 0 other;

Alignment Scores:
Pred. No.: 1.74e-13 Length: 4005
Score: 282.00 Matches: 86
Percent Similarity: 44.41% Conservative: 41
Best Local Similarity: 30.07% Mismatches: 123
Query Match: 14.06% Indels: 36
DB: Gaps: 24

US-10-007-527A-2 (1-379) x AAS17120 (1-4005)
QY 60 AenglyProLysGlySerglyPheGlyGlyLeuArgSerglyLysGlyTyrIleCys 79
Db 3636 GATGCGGGCGGCTCGGCGCATCTCTTTCAGCTTACGACAGTCCGCAATGTGGGGCTGC 3577
QY 80 ProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGlnIleSerglyValAla 99
Db 3576 GCAGTTTGTAGGCTCGCATGCGCAAAATTCGCGCTTGTGAATGAATCACTGTGCGG 3517
QY 100 HisGlnLeuGlyThrGlySerglyAlaMetValThrMetThrMetArgHisThrAla 118
Db 3516 TGGGGCGCGCAACAGGCTTTGTGCGGCTGCTCAATCTTCACGCCGACGACAAAGCA 3457
QY 119 GlyGlnArgLeuHisAspLeuTyrThrGlyLeuSerAlaIleTyrLeuAlaIleThrAsn 138
Db 3456 GGTGATTCGCTTTTGTGATCTTCTCCAAACATGAAGAGCGAACAAGGGTTCGTCAG 3397
QY 139 GlyArgGTPArgThrGluArgGlnMetTyrGlyCysAspGlyTyrValArgAlaVal 158
Db 3396 CGCCGCGAGTGGCGGATTAACCATTTGT-----GGGTCACTCACTCACT 3349
QY 159 GluIleThrHisGly---LysAsnGlyTyrPheHisValHisValHisAlaLeu---LeuMet 176
Db 3348 GAAATTACTCATAGCTAATGCAATGAGTGGACCGCGCACTTCAAGAAATTCCTTGTGCA 3289
QY 177 PheSerglyAspValSergly---AsnIleLeuGlnSergPheSeraSpAlaMetPhe 194
Db 3288 AGAGCTGGCGATGGAAGCGAAGCGCTGATGATGACGAACGCTACGCGATCG----- 3235
QY 195 AsparGTPThrSerLysLeuValSerLeuGlyPheAlaProLeuArgAsnSergly 214
Db 3234 -----TGGCGTGCATGTCGAAAGGAGTATGCGATGTGGGCGC-----AACGATGCG 3190
QY 215 GlyLeuAspValArgLysIleGlyGlyGluAlaAspGlnValLeuAlaAlaTyrLeuThr 234
Db 3189 GCGTTTGACGTTCGGGGGCTGCCAATGCGGGCGAC-----TAGGTAGCG 3145
QY 235 LysIleAlaSerGlyValGlyMetGluValGlySerglyAspGlyLysSerglyArgHis 254
Db 3144 AAG-----TGGGGCGCTGCCGCAAGAGTTCGCTTCAAGCTCGAAGAGCGGAAACCG 3091
QY 255 GlyAsnArgAlaProTyrGlnIleAlaValAspAlaValGlyLysAspProGlnAlaLeu 274
Db 3090 AAGAGCGACGCCAAGGACGACTACTT-----CAGCGCGCGGATGAC----- 3049
```

```
Qy 275 GluLeuTTPArgGluPheGluPheGlySerMetGly-----ArgArgAlaIleAlaTTP 292
Db 3048 GGGCTTTGGCTCGCAATATTTCATGCAACGAGCGGCAACGCGCAATTGGTGTGG 2989
Qy 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGluLeuThrAsp----- 308
Db 2988 AGCCCAAGGCTTGAAGAAGAAATGCGGTTTG--GTCGAGCTGGATGACGATGAAAGCTATG 2932
Qy 309 AlaGluIleValGluGluGluSerAlaProValAlaIleIleProAlaArg 328
Db 2931 GCTGAAGTTGATGCAACGACGCAAGCGGCTTGAATTTGTCCGGAAGTGGGACAAATGAA 2872
Qy 329 SerTPMetMetIleArg 334
Db 2871 GGGTGAAGCAAGTGC GC 2854
```

Search completed: August 22, 2003, 10:45:12  
Job time : 307 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: August 22, 2003, 10:35:44 ; Search time 224 Seconds  
(without alignments)  
3803.813 Million cell updates/sec

Title: US-10-007-527A-2  
Perfect score: 2005  
Sequence: 1 MTSVAEHSHSGKDRPVLVLS.....HHYHPLADVAPPIISVRK 379

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US10007527/runat\_20082003\_165152\_15399/app\_query.fasta\_1.519  
-DB=Published Applications NA -OFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPT=0 -LOOPT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human0.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAVISIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10007527 @CGN 1.164 @runat\_20082003\_165152\_15399  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEOUTPUT -NEG\_SCORES=0 -WAIT -DSBIOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FAPOP=6 -FAPEXT=7 -YCAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2005	100.0	1140	14	US-10-007-527A-1

2	2005	100.0	1140	14	US-10-007-527A-1	Sequence 1, Appli
3	2005	100.0	6334	14	US-10-007-527A-5	Sequence 5, Appli
4	2005	100.0	6334	14	US-10-007-452-5	Sequence 5, Appli
5	2005	100.0	9652	14	US-10-007-527A-7	Sequence 7, Appli
6	2005	100.0	9652	14	US-10-007-452-7	Sequence 7, Appli
7	2005	100.0	11241	14	US-10-007-527A-6	Sequence 6, Appli
8	2005	100.0	11241	14	US-10-007-452-6	Sequence 6, Appli
9	326	16.3	4447	9	US-09-835-381-5	Sequence 5, Appli
10	326	16.3	4447	14	US-10-136-232-5	Sequence 5, Appli
11	318	15.9	4447	9	US-09-835-381-7	Sequence 7, Appli
12	318	15.9	4447	14	US-10-136-232-6	Sequence 6, Appli
13	282	14.1	2401	9	US-09-826-191-6	Sequence 6, Appli
14	282	14.1	2401	14	US-10-263-666-6	Sequence 6, Appli
15	282	14.1	2401	14	US-10-261-481-6	Sequence 6, Appli
16	282	14.1	2401	14	US-10-261-481-6	Sequence 6, Appli
17	282	14.1	4005	9	US-09-826-191-2	Sequence 2, Appli
18	282	14.1	4005	14	US-10-263-666-2	Sequence 2, Appli
19	282	14.1	4005	14	US-10-261-481-2	Sequence 2, Appli
20	282	14.1	4005	14	US-10-261-481-2	Sequence 2, Appli
21	138	6.3	902568	14	US-10-156-761-1	Sequence 1, Appli
22	122	6.1	2871	9	US-09-815-242-7827	Sequence 7827, Ap
23	117	5.8	68750	13	US-10-014-717-1	Sequence 1, Appli
24	116	5.8	1542	9	US-09-833-745-7	Sequence 7, Appli
25	113	5.6	902568	14	US-10-156-761-1	Sequence 1, Appli
26	112.5	5.5	1326	9	US-09-833-745-12	Sequence 12, Appli
27	109.5	5.5	2715	14	US-10-156-761-2245	Sequence 2245, Ap
28	109.5	5.5	13842	10	US-09-861-289-30	Sequence 30, Appli
29	107	5.3	13842	11	US-09-861-289-30	Sequence 30, Appli
30	107	5.3	13842	11	US-09-988-384B-30	Sequence 30, Appli
31	107	5.3	13842	11	US-09-988-384B-30	Sequence 30, Appli
32	107	5.3	13842	11	US-09-861-289-5	Sequence 5, Appli
33	107	5.3	36778	10	US-09-861-289-5	Sequence 5, Appli
34	107	5.3	36778	10	US-09-861-289-5	Sequence 5, Appli
35	107	5.3	36778	11	US-09-836-821-5	Sequence 5, Appli
36	107	5.3	37948	11	US-09-988-384B-5	Sequence 5, Appli
37	107	5.3	38506	11	US-09-793-708-19	Sequence 19, Appli
38	107	5.3	38506	12	US-10-201-365-1	Sequence 1, Appli
39	106.5	5.3	15819	10	US-09-764-877-2665	Sequence 2665, Ap
40	106.5	5.3	23429	12	US-10-017-181-2003	Sequence 2, Appli
41	104.5	5.2	3687	11	US-09-917-363-2	Sequence 2, Appli
42	104.5	5.2	3687	11	US-09-917-363-2	Sequence 2, Appli
43	104.5	5.2	17410	8	US-08-957-425-3	Sequence 3, Appli
44	103.5	5.2	1251	14	US-10-156-761-2248	Sequence 3248, Ap
45	103.5	5.2	1860	9	US-09-815-242-4156	Sequence 4156, Ap

#### ALIGNMENTS

RESULT 1  
US-10-007-527A-1  
; Sequence 1, Application US/10007527A  
; Publication No. US20030044807A1  
; GENERAL INFORMATION:  
; APPLICANT: Brannucci, Jean-Francois  
; APPLICANT: Brannucci, Michael G.  
; APPLICANT: Chen, Qiong  
; APPLICANT: Koshichka, Kristy N.  
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
; FILE REFERENCE: C14709 US NA  
; CURRENT APPLICATION NUMBER: US/10/007,527A  
; PRIOR APPLICATION NUMBER: 2001-12-05  
; PRIOR FILING DATE: 2000-12-12  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 1  
; LENGTH: 1140  
; TYPE: DNA  
; ORGANISM: Rhodococcus AN12  
US-10-007-527A-1  
Alignment Scores: 1.08e-213 Length: 1140  
Pred. No.: 1

Score: 2005.00 Matches: 379  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 14 Gaps: 0

US-10-007-527A-2 (1-379) x US-10-007-527A-1 (1-1140)

QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20  
 Db 1 ATGACCAGCGTAAAGTCTGAACACCTTCCGGCAAAAGCCGGCTCCCTCGTGTG 60  
 QY 21 SerAspLysArgGlyTLeaArgHisLeuLeuArgProLysLeuGlnGlnLeThrThrSer 40  
 Db 61 TCCGATTAAAGCGCGCATCCGACACGAACCTGACCAAACTTCAACAAATCACACGTC 120  
 QY 41 GluThrPheAsnAlaCysGlyArgProLieserGlyValAsnGlyValThrIleValAsn 60  
 Db 121 GAAACATTTAAAGCTGTCGAGCCGCGCCGATTCCTGCGGTAAACGGGTGACATTCAC 180  
 QY 61 GlyProLysGlySerGlyPheGlyLeuArgSerCysGlyLysGlyTyrIleCysPro 80  
 Db 181 GGTTCGAAAGCTTCGTGATTCGAGGCTTCGTTCTGCGGAAAGGCTGATCTGCCCC 240  
 QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluLieserGlnValAlaHis 100  
 Db 241 TCGTGTGCGGAAAAAGTCGTCACATCGTCGACAGAAATTTCTCAAGTTGTGCTCAT 300  
 QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120  
 Db 301 CAACCTGGAGCTGAGATCTGTGCGATGTGACGATGACATCGCCATACAGCTGTGACG 360  
 QY 121 ArgLeuHisAspLeuThrThrGlyLeuSerAlaAlaTyrLysAlaAlaThrAsnGlyArg 140  
 Db 361 CGGCTCCAGCACTATGACCTGACCTTTCGCGACCTCGAAAGCTCCGACCAACGCTCT 420  
 QY 141 ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160  
 Db 421 CGTTGGCGTACGGAACGTGAATGTACGCTGCGACGAGATACGTCGCGCTGTGAAATC 480  
 QY 161 ThrHisGlyLysAsnGlyTyrPheIleValHisValHisAlaLeuLeuMetPheSerGlyAsp 180  
 Db 481 ACTCAACGAAAAAAGCGCTGGACGCTCAAGTTCAGCGCTCACTGTCAGTGTGTGAC 540  
 QY 181 ValSerGluAsnIleLeuGlnSerPheSerAspAlaMetPheAspArgTyrThrSerLys 200  
 Db 541 GTAGAGGAGAACATCTCTGATCTCTCTCGATGCAATGTCATCGCGACCTTCCAA 600  
 QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysLeuAspValArgLys 220  
 Db 601 CTCGATCTCTGGGATTTTCTGCGGCACATACGTAATTCGGGTGTCTCGATGTACGAAAG 660  
 QY 221 IleGlyGlyGluAlaAspGlnValIleuAlaAlaTyrLeuThrIleIleAlaSerGlyVal 240  
 Db 661 ATCGGGGGTGAAGCTATACATGCTCGCGCTATCTGACGAAATTCATCTGCGCT 720  
 QY 241 GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyLysAsnArgAlaProTyr 260  
 Db 721 GGTATGAGAGTTGTGTGTGCGCGGCGGAAAAAGTGTGTGACATGTGCAACCTGTG 780  
 QY 261 GluIleAlaValAspAlaValGlyLysAspProGlnAlaLeuGlnLeuThrArgGluPhe 280  
 Db 781 GAAATCGCTGTATGATCAGTGGCGCGGAGTCCCAAGCCGTGGAACTGTGCGCAAGATT 840  
 QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAlaAspGly 300  
 Db 841 GAGTTTGTTCGTGATGGAGCGTCGGGCAATGCGCTGTCTCCGTGAGATTGCGTCCAGCT 900  
 QY 301 GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGlnGlnGlnLeuSerAlaProVal 320  
 Db 901 GGTCTTGGGGCAGAACTTAACAGATGCTCAGATGTTGAGAGGAAAGATCTGCCCGGCT 960  
 QY 321 MetValAlaIleIleProAlaArgSerTyrMetMetIleArgThrCysAlaProTyrVal 340

Db 961 ATGTGTGGATCATTCGGCGCATCTGATGATGATGATTCGACTTGTGCGCTTACGTC 1020  
 QY 341 PheGlyGluIleLeuGlyLeuValGlyAlaValThrTyrPheGlyAsnLeuArgAspHis 360  
 Db 1021 TTCGGCGAGATCTCTCGACTGTGTGAACTGGCGGACTTGGGAAATCTTCGTGATCAC 1080  
 QY 361 LeuHisTyrArgLeuProAlaAlaAspValArgProLieserValArgLys 379  
 Db 1081 TTGCAATTATCATTTGCCCGGACCGGATGTGCGGCCCCGATATATATCGTTCCGAAG 1137  
 RESURF 2  
 US-10-007-452-1  
 ; Sequence 1, Application US/10007452  
 ; Publication No. US20030093701A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tomb, Jean-Francois  
 ; APPLICANT: Brumet, Michael G.  
 ; APPLICANT: Chen, Qiong  
 ; APPLICANT: Kostichka, Kristy N.  
 ; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
 ; FILE REFERENCE: CL1709 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/007,452  
 ; PRIOR FILING DATE: 2001-11-08  
 ; PRIOR APPLICATION NUMBER: 60/254,868  
 ; PRIOR FILING DATE: 2000-12-12  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 1  
 ; LENGTH: 1140  
 ; TYPE: DNA  
 ; ORGANISM: Rhodococcus AM12  
 ; US-10-007-452-1  
 Alignment Scores:  
 Pred. No.: 1,08e-213 Length: 1140  
 Score: 2005.00 Matches: 379  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 14 Gaps: 0  
 US-10-007-527A-2 (1-379) x US-10-007-452-1 (1-1140)  
 QY 1 MetThrSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20  
 Db 1 ATGACCAGCGTAAAGTCTGAACACCTTCCGGCAAAAGCCGGCTCCCTCGTGTG 60  
 QY 21 SerAspLysArgGlyTLeaArgHisLeuLeuArgProLysLeuGlnGlnLeThrThrSer 40  
 Db 61 TCCGATTAAAGCGCGCATCCGACACGAACCTGACCAAACTTCAACAAATCACACGTC 120  
 QY 41 GluThrPheAsnAlaCysGlyArgProLieserGlyValAsnGlyValThrIleValAsn 60  
 Db 121 GAAACATTTAAAGCTGTCGAGCCGCGCCGATTCCTGCGGTAAACGGGTGACATTCAC 180  
 QY 61 GlyProLysGlySerGlyPheGlyLeuArgSerCysGlyLysGlyTyrIleCysPro 80  
 Db 181 GGTTCGAAAGCTTCGTGATTCGAGGCTTCGTTCTGCGGAAAGGCTGATCTGCCCC 240  
 QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluLieserGlnValAlaHis 100  
 Db 241 TCGTGTGCGGAAAAAGTCGTCACATCGTCGACAGAAATTTCTCAAGTTGTGCTCAT 300  
 QY 101 GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120  
 Db 301 CAACCTGGAGCTGAGATCTGTGCGATGTGACGATGACATCGCCATACAGCTGTGACG 360  
 QY 121 ArgLeuHisAspLeuThrThrGlyLeuSerAlaAlaTyrLysAlaAlaThrAsnGlyArg 140  
 Db 361 CGGCTCCAGCACTATGACCTGACCTTTCGCGACCTCGAAAGCTCCGACCAACGCTCT 420  
 QY 141 ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160

```

Db      421 CGTTGGCGTACGGAAGTGAAATGTAACGCTGGCAGGAACTGCGCGCTGTGAAATC 480
QY      161 ThrHISGlyLYsaSngLYTrpHISValHISValHISaleuLeuMetPheSerGlyAASP 180
Db      481 ACTCAGGAAAAAAGCGCTGGCAGCTCAAGTTCAACGCTTCACTATGTTCAATGGTGAC 540
QY      181 ValSerGluAsnLLeuGluSerPheSerAspAlaMetCysAspArgTrpThrSerLYS 200
Db      541 GTAGTGAAGAACATCTCGAATCTTCTCGAAGCAATGTCATGGTGAAGCTTCCAAA 600
QY      201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLYLeuAspValArgLYS 220
Db      601 CTCGTATCTCTGGGATTTGCTGGCCACACGTAATTCGGTGTCTCGATGTACAAAG 660
QY      221 ILeGlyGlyGluAlaAspGluValLeuAlaAlaTYrLeuThrLYSILEaSerGlyVal 240
Db      661 ATCGGCGGTGAAGCTGATCAAGTTCTCGCTGCTATCTGACGAAATTCATCTGGCGTT 720
QY      241 GlyMetGluValGlySerGlyAspGlyLYSserGlyArgHISGlyAsnArgAlaProTrp 260
Db      721 GGTATGAGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780
QY      261 GluIleAlaValAspAlaValGlyLYSAspProGluAlaLeuGluLeuTrpArgGlyLys 280
Db      781 GAATTCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
QY      281 GluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAla 300
Db      841 GAGTTTGTTTCATGTGGAGCGTGGCGCAATCGCGTGTCTCCGTGTGTGTGTGTGTGTGT 900
QY      301 GlyLeuGlyValGluLeuThrAspAlaGluIleValGluGluGluGluSerAlaProVal 320
Db      901 GGTCTGGGCGCAGAACTAAACAGATGCTCAGATGCTGAGCAGAGAAATCTGCCCGGTC 960
QY      321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTYrVal 340
Db      961 ATGGTTGCAATCTTCGCGCGCATGCTGATGATGATGATGATGATGATGATGATGATGAT 1020
QY      341 PheGlyGluIleLeuGlyLeuValGluAlaGlyAlaThrTrpGluAsnLeuArgAspHIS 360
Db      1021 TTCGGGGAATCTCTCGAATCTCTCGAAGCTGGCGCACTTGGGAAATCTTGTATCTAC 1080
QY      361 LeuHISLYrArgLeuProAlaAlaAspValArgProProIleIleSerValArgLYS 379
Db      1081 TTGCATTATCGATTGCCCGCAGCGGATGTGGCGCCCGCATATATCGCTTGCAGAG 1137

RESULT 3
US-10-007-527a-5/c
/ Sequence 5, Application US/10007527A
/ Publication No. US20030044807A1
/ GENERAL INFORMATION:
/ APPLICANT: Tomb, Jean-Francois
/ APPLICANT: Bramucci, Michael G.
/ APPLICANT: Cheng, Qiong
/ APPLICANT: Kostichka, Kristy N.
/ TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
/ CURRENT APPLICATION NUMBER: US/10/007,527A
/ PRIOR FILING DATE: 2001-12-05
/ PRIOR APPLICATION NUMBER: 60/254,868
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: Microsoft Office 97
/ SEQ ID NO 5
/ LENGTH: 6334
/ TYPE: DNA
/ ORGANISM: Rhodococcus AN12
US-10-007-527a-5

Alignment Scores:
Pred. No.: 1 02e-212 Length: 6334
Score: 2005.00 Matches: 379

```

```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
US-10-007-527a-2 (1-379) x US-10-007-527a-5 (1-6334)

QY      1 MetThrSerValSerAlaGluHISLeuSerGlyLYSAspArgProProValLeuValSer 20
Db      3051 ATGACCAAGCGTAAGTCTGTAACACTTTCGGCAAAAGCCGGCCTCCGCTCGTGTG 2992
QY      21 SarAspLYrArgGlyLYLeuArgHISGluLeuAspProLYSLeuGluIleThrThrSer 40
Db      2991 TCGATTAACCGCGGATTCGGAGCAACGCCAACCCCAACTTCAACAAATCAACAGTCA 2932
QY      41 GluThrPheAsnAlaCysGlyYrArgProIleSerGlyValAsnGlyValThrIleValAsn 60
Db      2931 GAAACATTAAAGCTCTGTGGCCGCGCATTTTGGCGGTAAACGGTGTACATTTGTCAAC 2872
QY      61 GlyProLYSGLYSerGlyPheGlyLYLeuArgSerCysGlyLYSGLYTrpIleCysPro 80
Db      2871 GGTCCGAAAGTTCTGATTCGAGGCGCTTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGT 2812
QY      81 CysCYeAlaGlyLYValGlyValAlaHISArgAlaAspGluIleSerGluValAlaHIS 100
Db      2811 TCTGTGCGGAAAAAGTCGTGCACATGTGTGAAGCAAAATTTCTCAAGTTGTGTCTCAT 2752
QY      101 GluLeuGlyThGlySerValAlaMetValThrMetThrMetArgHISThrAlaGlyGln 120
Db      2751 CAATCGGAGCTGAGATCTGTTCGATGTGAAGATGATGATGATGATGATGATGATGATGAT 2692
QY      121 ArgLeuHISAspLeuTrpThrGlyLeuSerAlaAlaTrpLYSAlaAlaThrAsnGlyArg 140
Db      2691 CGGCTCCAGCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2632
QY      141 ArgTrpArgThrGluArgGluMetTYrGlyCysAspGlyTYrValArgAlaValGluIle 160
Db      2631 CGTTGCGGTACGGAACGTGAAATGTACCGCTCGACGAGTACGTGGCGCTGTGTAAATC 2572
QY      161 ThrHISGlyLYsaSngLYTrpHISValHISValHISaleuLeuMetPheSerGlyAASP 180
Db      2571 ACTCAGGAAAAAAGCGCTGGCAGCTGACGCTCACGCTCACTCATATGTTGATGTGTAC 2512
QY      181 ValSerGluAsnLLeuGluSerPheSerAspAlaMetPheAspArgTrpThrSerLYS 200
Db      2511 GGAATGGAACATCTCTCGAATCTCTCTCGATGATGATGATGATGATGATGATGATGATGAT 2452
QY      201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLYLeuAspValArgLYS 220
Db      2451 CTCGTATCTCTGGGATTTGCTGGCCACACGTAATTCGGTGTGTGTGTGTGTGTGTGTGT 2392
QY      221 ILeGlyGlyGluAlaAspGluValLeuAlaAlaTYrLeuThrLYSILEaSerGlyVal 240
Db      2391 ATCGGCGGTGAAGCTGATCAAGTTCTCTGTGGTATCTGACGAAATTCATCTGCGCTT 2332
QY      241 GlyMetGluValGlySerGlyAspGlyLYSserGlyArgHISGlyAsnArgAlaProTrp 260
Db      2331 GGTATGAGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2272
QY      261 GluIleAlaValAspAlaValGlyLYSAspProGluAlaLeuGluLeuTrpArgGlyLys 280
Db      2271 GAATTCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2212
QY      281 GluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAla 300
Db      2211 GAGTTTGTTTCATGTGAAGCGTGGCAATCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2152
QY      301 GlyLeuGlyValGluLeuThrAspAlaGluIleValGluGluGluGluSerAlaProVal 320
Db      2151 GGTCTTGGGCGCAAGCTTAACAGATCTCAGATGCTTGAACAGAAAGAAATCTCCCGGTC 2092
QY      321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTYrVal 340

```

```
Db      2091  ATGGTTGCATATTCGGCGCGGCGATCGATGATGATTCGGATCTTGCGCCTTAACGTC 2032
QY      341  PhegiyguilleleuuglyleuvalaglyalathrrpghuasnuargaspHis 360
Db      2031  TTCGGGAGATCCTCGGATCTCGATCGAGCTGGGGGAGACTTGGGAAATCTTGATGATC 1972
QY      361  LeuHsiTyrrgleuProAlaAlaaspValargProProllelleservAlarglys 379
Db      1971  TTGCATTAATCGATTGCCCGCAGCGGATGCGCGCCCGCATATAATCGGTTGCCAAG 1915

RESULT 4
US-10-007-452-5/c
; Sequence 5, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 5
; LENGTH: 6334
; TYPE: DNA
; ORGANISM: Rhodococcus AN12
US-10-007-452-5

Alignment Scores:
Pred. No.: 1,02e-212      Length: 6334
Score: 2005.00      Matches: 379
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 14      Gaps: 0

US-10-007-527a-2 (1-379) x US-10-007-452-5 (1-6334)
QY      1  MetThrSerValSerAlaGluHisLeuSerGlyValAspArgProProValleuValSer 20
Db      3051  ATGACCAAGCTAAGTCTGTAACACCTTTCGGCGAAAGACCGGCGCTCCGCTCCGATGCG 2992
QY      21  SerAspLyArgGlyIleArgHisGluLeuArgProLyLeuGlnGlnIleThrThrSer 40
Db      2991  TCCGATTAACCGCGGATCCGGCGACGAACTGCCGCCCAACTTCACAAATCCACGACGTCA 2932
QY      41  GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
Db      2931  GAAACATTAAACCCCTGTGGCCGCGCGGATTTCTGGCGTAAACGCTGTACCATTTGCAAC 2872
QY      61  GlyProLySerGlyPheGlyGlyLeuArgSerCysGlyValysGlyTyrPileCysPro 80
Db      2871  GGTCCGAAAGTTCTCGATTCGAGGCGCTTCGTTCTCGCGAAAGGGCTGATCTGCCCC 2812
QY      81  CysCysAlaGlyValValAlaHisArgAlaAspGluIleSerGlnValAlaHis 100
Db      2811  TGTGTGCGGGGAAATCGGTGCACATCGGACGAGAAATTTCTAAAGTTGTGTCTCAT 2752
QY      101  GlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
Db      2751  CAACCTGGGAGCTGATCTGTGGCGATGCGATGACATACATGCGCATACAGCTGCTCAG 2692
QY      121  ArgLeuHisAspLeuThrThrGlyLeuSerAlaAlaTyrPileysAlaIleThrAsnGlyArg 140
Db      2691  CGGCTCCACGACCTATGACATGACATTTCCGACGCTCGGAAAGCTCGACCAACGAGTCT 2632
QY      141  ArgTyrArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle 160
```

```
Db      2631  CGTTGGCGTACGGAACGTGAATGTACAGGCTTGCAAGGATACGTGCCGCTGTGGAATC 2572
QY      161  ThrHisGlyLysAsnGlyTyrPheIleValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
Db      2571  ACTCACGGGAAAAAAGCGCTGGCACGTCACAGTTCCACGGGCTACTCATGTCAGTGTAC 2512
QY      181  ValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTyrThrSerLys 200
Db      2511  GTAGTGTGAACATCCCTCGAATCCCTTCGCGATGCGATGTTGATCGGTGTGACATTCCAA 2452
QY      201  LeuValSerLeuGlyPheAlaAlaProleuArgAsnSerGlyValLeuAspValArglys 220
Db      2451  CTCGATATCTCGGATTTGCTGCCCATTAAGTATTCGGGTGATCTCGATGTACGAAAG 2392
QY      221  IleGlyGlyGluAlaAspGlnValleuAlaAlaTyrLeuThrLysIleAlaSerGlyVal 240
Db      2391  ATCCGCGGTGAAGCTGACCAAGTTCTGCTCGTATTCGAGAAATTCGATCTGGCGTT 2332
QY      241  GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTyr 260
Db      2331  GGTATGAGGTTGGTGTGTGGCGACGGAAGAGTGTGCACATGGCACCGTGCACCTTGG 2272
QY      261  GluIleAlaValAlaAspAlaValGlyValAspProGlnAlaLeuGluLeuThrArgGluPhe 280
Db      2271  GAATTCGCTGTGATGACGTGGCGCGGATTCACAAAGCTTGGAACTGTGGCGAATTT 2212
QY      281  GluPheGlySerMetGlyArgArgAlaIleAlaTyrSerArgGlyLeuArgAlaArgAla 300
Db      2211  GAGTTTGGTTCGATGGACGTGCGGCAATCCGATGTCCTCCGATGTCGTCGCCGAGCT 2152
QY      301  GlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluGlnSerAlaProVal 320
Db      2151  GGTCTTGGGCGAAGACTAACAAGTCTCAGATCGTTGACGAGAAATCTGCCCGGTC 2092
QY      321  MetValAlaIleIleProAlaArgSerTrpMetIleArgThrCysAlaProTyrVal 340
Db      2091  ATGGTTGGAGATATTCGCGCGGATCGAGATGATGATTCGACCTTGCGCCTTAACGTC 2032
QY      341  PhegiyguilleleuuglyleuvalaglyalathrrpghuasnuargaspHis 360
Db      2031  TTCGGGAGATCCTCGGATCTCGATCGAGCTGGCGCGGACTTGGGAAATCTTGATGATC 1972
QY      361  LeuHsiTyrrgleuProAlaAlaaspValargProProllelleservAlarglys 379
Db      1971  TTGCATTAATCGATTGCCCGCAGCGGATGCGCGCCCGCATATAATCGGTTGCCAAG 1915

RESULT 5
US-10-007-527a-7/c
; Sequence 7, Application US/10007527a
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527a
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; TYPE: DNA
; ORGANISM: Plasmid pRHB17
US-10-007-527a-7

Alignment Scores:
Pred. No.: 1,76e-212      Length: 9652
Score: 2005.00      Matches: 379
Percent Similarity: 100.00%      Conservative: 0
```



Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 14 Gaps: 0

US-10-007-527a-2 (1-379) x US-10-007-527a-7 (1-9652)

```
QY 1 MetThSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
DB 5052 ATGACCGAGGTAAAGTGTGAACACCTTTCGGCAAAACGGCCCTCCGCTCCGTGTCAG 4993

QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40
DB 4992 TCCGATTAAGCGCGGCAATCCGCGACGAATCGACCAATTTCAACAAATCACACGTCAC 4933

QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
DB 4932 GAAACATTTTAACCGCTGTGCGCGGCAATTTCTGCGGTGAACGGTGTGACCATGTGTCAAC 4873

QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerGlyLysGlyYTPPIleCysPro 80
DB 4872 GGTCCGAAAGGTTCTGGATTGCGAGGCGCTTCGTTCCGCGAAAGGGCTGATCTGCCCC 4813

QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHis 100
DB 4812 TGCCTGCGCGGAAAGTCGCTGCAATCGTCAGACGAATTTCTCAAGTTGTGCTCAT 4753

QY 101 GlnLeuGlyYThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
DB 4752 CAACCTCGGAGCTGGATCTGTGCGATGTGACATGACATGCGCCATACAGCTGTGTGACG 4693

QY 121 ArgLeuHisAspLeuThrPheGlyLeuSerAlaIleThrIleValAlaHisArg 140
DB 4692 CGGCTCCACGACCTATGTGATGTGACATTTTCGCGACGCCCGGAAAGCTGTGCAACCGGTCTG 4633

QY 141 ArgTrpArgThrGlnArgGluMetYrGlyCysAspGlyYrValAlaArgAlaValGluIle 160
DB 4632 CGTTGGCGTACGGAACGTGAATGTACGCTGCGACGAGATACGTGCGCGCTGTGAAATTC 4573

QY 161 ThrHisGlyLysAsnGlyYrPheHisValHisValHisAlaLeuLeuMetPheSerGlyAsp 180
DB 4572 ACTCACGGAAGAAACCGCTGCGACGTGCAAGTTCAAGCGCTACATGTTCAAGTGTGAC 4513

QY 181 ValSerGluAsnGlnIleLeuGluSerPheSerAspAlaMetPheAspArgTrpThrSerLys 200
DB 4512 GTAGGTGAAGAACATCTCGAATCTCTTCGATGCGATGCTTCATCGTGAGCTTCCAA 4453

QY 201 LeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLys 220
DB 4452 CTGATCTCTGGGATTTGCTGCGCCACTACGTAATTCGGGTGTCTCGATGTACGAAG 4393

QY 221 IleGlyGlyValAlaAspGlnValLeuAlaAlaYrLeuThrLysIleAlaSerGlyVal 240
DB 4392 ATCGGCGGTGAAGCTGATCAAGTTCTCGCTGCGGATCTGACGAAATTCATCTGGGTT 4333

QY 241 GlyMetGluValGlySerGlyArgGlyLysSerGlyYrArgHisGlyValAsnAlaArgTrp 260
DB 4332 GGATATGAAGGTGTGTATGTGCGACGCGAAAGTGTGTGACATGTGCAACCTGTGACCTGG 4273

QY 261 GluIleAlaValAlaAspAlaValGlyYrAspProGlnAlaLeuGlnLeuThrPArgGluPhe 280
DB 4272 GAAATCCCTTTGATGTGAGTGGGCGGGAATCCACAAACGTTGGACTGTGTGCGAATTT 4213

QY 281 GluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAla 300
DB 4212 GAGTTGTGTTGATGTGAGACGTGCGGCAATGCGGTGTGCTCGGTGATTTGGTGTGCGGAGCT 4153

QY 301 GlyLeuGlyValAlaGluLeuThrAspAlaGlnIleValGlnGlnGluSerAlaProVal 320
DB 4152 GGTCTTGGGCGCAAGATCAAGATGTCTCAATGTGTGACGAGAAAGATCTGCCCCGCTC 4093

QY 321 MetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProYrVal 340
DB 4092 ATGTGTGCGATCAATCCGCGCGCATGTGATGTATGTGACTGTGTGCGCTTGTACGTC 4033
```

```
QY 341 PheGlyGlnIleLeuGlyLeuValGluAlaGlyAlaThrTrpGluAsnLeuArgAspHis 360
DB 4032 TTCCGGAGAGATCTCTCGACCTGCGAAGCTGGCGCACTTGGGAAATCTTGTGTGATCAC 3973

QY 361 LeuHisThrArgLeuProAlaAlaAspValArgProProIleIleSerValArgLys 379
DB 3972 TTGCATTATCGAATGTGCGCGAGGATGTGCGGCCCCCGCATATATATGATTCGACAG 3916
```

## RESULT 6

```
US-10-007-452-7/c
; Sequence 7, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Olong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCES: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 9652
; TYPE: DNA
; ORGANISM: Plasmid pRHB17
US-10-007-452-7
```

## Alignment Scores:

```
Pred. No.: 1,766-212 Length: 9652
Score: 2005.00 Matches: 379
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
```

US-10-007-527a-2 (1-379) x US-10-007-452-7 (1-9652)

```
QY 1 MetThSerValSerAlaGluHisLeuSerGlyLysAspArgProProValLeuValSer 20
DB 5052 ATGACCGAGGTAAAGTGTGAACACCTTTCGGCAAAACGGCCCTCCGCTCCGTGTCAG 4993

QY 21 SerAspLysArgGlyIleArgHisGluLeuArgProLysLeuGlnGlnIleThrThrSer 40
DB 4992 TCCGATTAAGCGCGGCAATCCGCGACGAATCGACCAATTTCAACAAATCACACGTCAC 4933

QY 41 GluThrPheAsnAlaCysGlyArgProIleSerGlyValAsnGlyValThrIleValAsn 60
DB 4932 GAAACATTTTAACCGCTGTGCGCGGCAATTTCTGCGGTGAACGGTGTGACCATGTGTCAAC 4873

QY 61 GlyProLysGlySerGlyPheGlyGlyLeuArgSerGlyLysGlyYTPPIleCysPro 80
DB 4872 GGTCCGAAAGGTTCTGGATTGCGAGGCGCTTCGTTCCGCGAAAGGGCTGATCTGCCCC 4813

QY 81 CysCysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHis 100
DB 4812 TGCCTGCGCGGAAAGTCGCTGCAATCGTCAGACGAATTTCTCAAGTTGTGCTCAT 4753

QY 101 GlnLeuGlyYThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln 120
DB 4752 CAACCTCGGAGCTGGATCTGTGCGATGTGACATGACATGCGCCATACAGCTGTGTGACG 4693

QY 121 ArgLeuHisAspLeuThrPheGlyLeuSerAlaIleThrIleValAlaHisArg 140
DB 4692 CGGCTCCACGACCTATGTGATGTGACATTTTCGCGACGCCCGGAAAGCTGTGCAACCGGTCTG 4633

QY 141 ArgTrpArgThrGlnArgGluMetYrGlyCysAspGlyYrValAlaArgAlaValGluIle 160
DB 4632 CGTTGGCGTACGGAACGTGAATGTACGCTGCGACGAGATACGTGTGCGCTGTGAAATTC 4573
```



QY	1	MetThrSerValSerAlaGluHsiSleuSerGlyLysAspArgProProValIleuValSer	20
Dp	6641	ATGACCAACCGTAAGTGTCAAAACCTTTTCCGGCAAAAGACCGGCTCCCGTCTCGTGTG	6588
QY	21	SerAspLysArgGlyIleArgHsiSgluDeuArgProLysSleuGlnGlnIleThrThrSer	40
Dp	6581	TCCGATAGACCGCGGCATCCGGACGAACTGGCGAACCTTAACCAATCAACCACTCA	6522
QY	41	GluThrPheAsnAlaCysGlyValArgProIleSerGlyValAsnGlyValThrlIeValAsn	60
Dp	6521	GAACATTTTAACGCTGTGGCGCGCGCAATTTCTGGCGGAACCGGTGTACCACTGTGTAC	6463
QY	61	GlyProLysGlySerGlyPheGlyValLeuArgSerCysGlyLysGlyTyrIleCysPro	80
Dp	6461	GCTCCGAAGGTTCTCGAATTCGGAGGCTTTCGTTCTTCGCGAAGGCTGGAATCTGCC	6400
QY	81	CysCysAlaGlyLysValGlyValAsiAsArgAlaAspGlnIleSerGlnValValAlaHis	100
Dp	6401	TGCTGTGCGGAAAAAGTCGTGCACATGTGTGACGAGAAATTTTCCAAGTTGTTGCTCAT	6344
QY	101	GluIleuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGln	120
Dp	6341	CAACTCGGAGCTGAGTCTGTTCGTCGATGGAGACAGACCAATGCGGCATACAGCTGTGAG	6288
QY	121	ArgLeuHsiAspLeuIleTyrThrGlyLeuSerAlaAlaTyrLysAlaAlaThrAsnGlyArg	140
Dp	6281	CGGCTCCACGACCTATGACATGGACATTTCCGGACCTCGAAAAGTCGCGACCAACGCTGCT	6222
QY	141	ArgTyrIleThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValGluIle	160
Dp	6221	CGTTGGCGTACGGAACGTGAATGTACCGCTCGACCGAATACGTGCGCGCTGTGTAATCT	6163

```

US-09-835-361-5
RESULT 9
; Sequence 5, Application US/09835381
; Patent No. US20020045223A1
; GENERAL INFORMATION:
; APPLICANT: SUGA, MikiKO
; APPLICANT: ASAKURA, Yoko
; APPLICANT: MORI, YukiKO
; APPLICANT: ITO, HiSAO
; APPLICANT: KURAHASHI, Osamu
; TITLE OF INVENTION: ARGinine REPRESSOR DEFICIENT STRAIN OF Corynebacterium ATCC 14605
; TITLE OF INVENTION: FOR PRODUCING L-Arginine
; FILE REFERENCE: 206018USO
; CURRENT APPLICATION NUMBER: US/09/835,381
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: JP2000-129167
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 4447
; TYPE: DNA
; ORGANISM: Brevibacterium lactofermentum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1318)..(2598)
; OTHER INFORMATION:
US-09-835-361-5

```

## Alignment Scores:

Pred. No.:	6,05e-26	Length:	4447
Percent Similarity:	326.00	Matches:	101
Best Local Similarity:	47.75%	Conservative:	58
Query Match:	30.33%	Mismatches:	133
	16.26%	Indels:	42
		Gaps:	14

US-10-007-527A-2 (1-379) x US-09-835-381-5 (1-4447)

```
QY      34  LeuGlnGlnIleThrSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
      1396  ATGTACAGATCAACCAATAGTAAAGCTCTGCGCGGGGCGCATGAGTG-GGCGAGGAGACA 1454
      54  Asn-----GlyValThrIleVal-AsnGlyProIlySerGlyPheGlyLeuArg 71
      1455  AGCTGTTCGGGTCCGTGGTCTAACGGT-----GCTTCGCAATTGAGGGTCTGCA 1508
QY      71  gSerCysGlyLysGlyTrrIleCysProCysSalaglyLysValGlyValahisArgAl 91
      1509  AAACCTCACTCGCGTGGGGGCGACCTCTCGCTGAATTGAAAGTCATGGCGGACCGCG 1568
QY      91  aAspGluIleSerGlnValValahisGlnLeuGlyThrGlySerValAlaMetValTh 111
      1569  CATTAAGCTGCTATTGCTTAAGTAAGATCACTTGGCGGCGGGCGGCGCTCATGATGTT 1628
QY      111  rMet---ThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTrrThrGlyLeuSe 130
      1629  TGTGGGCACTGTTGACACACACCGCTCAAGTCATTTGCCAGAGTTGAAGCGGATATTA 1688
QY      130  rAlaAlaTrrPlySalAlaAlaThrAsnGlyArgArgTrrPArgThrGluArgGluMetYrG 150
      1689  GACTGGCTACTCTTCGATGTTGAAGAAACATCTCACTGGAGAAAGAACGTGACAGGTACG 1748
QY      150  yCysAspGlyTrrValArgAlaValGluIleThrHisGly---LysAsnGlyTrrHisVa 169
      1749  GGTGGAGACACACCTATAGTACTATGAGTCAAGTCAACAGACTCTTGGGCGGAGGTTGG 1808
QY      169  IHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
      1809  GACCGGACACATGCTGTGTGTTCTTGATCGCTCACTGTTCAGACATCAACTCAAGGGCTT 1868
QY      189  eSerAspAlaMetPheAspArgTrrThrSerIlySleValSerLeuGlyPheAlaAlaPr 209
      1869  TGAAGATTCATGTTTCCCGCTGCTGCTCGGTGTTAAAGCCGGTATGACCGCGCC 1928
QY      209  oLeuArgAsnSerGly---GlyLeuAsp---ValArgLysIleGlyGlyValAlaAspG 227
      1929  ACTGCGTGAACACGCGGTCAAACTTGATCAGGTCTTCACTGAGGTGAGACGCTGCGAA 1988
QY      227  nValLeuAlaAlaTrrLeuThrIlySleAlaSerGlyValGlyMetGlyValGlySerG 247
      1989  A---ATGGCAACCTACTCCGCTAG---GGCATGTCTCAGAACTGACTGAGCTC 2036
QY      247  yAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrrPqluIleAlaValAspAlaVa 267
      2037  CGCTACACAAACCGCGCTAAGCGGTGCTACACCGCTTCAGATG---TTGATATGTT 2093
QY      267  lGlyGlyAspProGlnAlaLeuGlu-----LeuTrrPArgI 279
      2094  GGCGGATCAAAAGCACCGCGCGGAGATATGACGCTGTTTGTGCTGCGTGGCGCTGA 2153
QY      279  uPheGluPheGlySerMetGlyArgArgAlaIleAlaTrrSerArgIlyLeuArgAlaAr 299
      2154  GTATGAGGTGTGTTCTAAAAAAGCTGCGTTG---TCTGTGCTCAGGTGGG-----GCTTA 2204
QY      299  gAlaGlyLeuGlyValaGluLeuThrAspAlaGlnIle-----ValGluGlnGlu- 315
      2205  GCGGCTTGGGCAATTGATTAACATACGCTGATGTCACCTGCAATGAAGTGAAGAAAGACT 2264
QY      316  -----GluSerAlaProValMetValAl 323
      2265  GTACAAAGCTCGCGGTCTGGAAGACGGAACGAGGTGAATCAACCCGCTGTGCTGTTC 2324
```

```
QY      323  alleIleProAlaArgSerTrrPmetMetIleArgThr 335
      2325  TTTGGTGAAGCCCGCATGATTTGGAAGACTATTCAGTCT 2361
```

## RESULT 10

```
US-10-196-232-5
; Sequence 5, Application US/10196232
; Publication No. US20030113899A1
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, MIKIKO
; APPLICANT: ITO, HISAO
; APPLICANT: GUNOI, YOSHITVA
; APPLICANT: YASUEDA, HISASHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-ARGININE
; FILE REFERENCE: 225391USO
; CURRENT APPLICATION NUMBER: US/10/196,232
; PRIORITY FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: JP 2001-224586
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 5
; LENGTH: 4447
; TYPE: DNA
; ORGANISM: Brevibacterium lactofermentum
US-10-196-232-5
```

## Alignment Scores:

Pred. No.:	6,05e-26	Length:	4447
Score:	326.00	Matches:	101
Percent Similarity:	47.75%	Conservative:	58
Best Local Similarity:	30.33%	Mismatches:	133
Query Match:	16.26%	Indels:	42
		Gaps:	14

US-10-007-527A-2 (1-379) x US-10-196-232-5 (1-4447)

```
QY      34  LeuGlnGlnIleThrSerGluThrPheAsnAlaCysGlyArgProIleSerGlyVal 53
      1396  ATGTACAGATCAACCAATAGTAAAGCTCTGCGCGGGTGCATAGTG-GGCGAGGAGACA 1454
      54  Asn-----GlyValThrIleVal-AsnGlyProIlySerGlyPheGlyLeuArg 71
      1455  AGCTGTTCGGGTCCGTGGTCTAACGGT-----GCTTCGCAATTGAGGGTCTGCA 1508
QY      71  gSerCysGlyLysGlyTrrIleCysProCysSalaglyLysValGlyValahisArgAl 91
      1509  AAACCTCACTCGCGTGGGGGCGACCTCTGCTGAATTGAAAGTCATGGCGGACCGCG 1568
QY      91  aAspGluIleSerGlnValValahisGlnLeuGlyThrGlySerValAlaMetValTh 111
      1569  CATTAAGCTGCTATTGCTTAAGTAAGATCACTTGGCGGCGGGCGGCGCTCATGATGTT 1628
QY      111  rMet---ThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTrrThrGlyLeuSe 130
      1629  TGTGGGCACTGTTGACACACACCGCTCAAGTCATTTGGCGAGATTTGAAGCGGATATTA 1688
QY      130  rAlaAlaTrrPlySalAlaAlaThrAsnGlyArgArgTrrPArgThrGluArgGluMetYrG 150
      1689  GACTGGCTACTCTTCGATGTTGAAGAAACATCTCACTGGAGAAAGAACGTGACAGGTACG 1748
QY      150  yCysAspGlyTrrValArgAlaValGluIleThrHisGly---LysAsnGlyTrrHisVa 169
      1749  GGTGGAGACACACCTATAGTACTATGAGTCAAGTCAACAGACTCTTGGGCGGAGGTTGG 1808
QY      169  IHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPh 189
      1809  GACCGGACACATGCTGTGTGTTCTTGATCGCTCACTGTTCAGACATCAACTCAAGGGCTT 1868
QY      189  eSerAspAlaMetPheAspArgTrrThrSerIlySleValSerLeuGlyPheAlaAlaPr 209
      1869  TGAAGATTCATGTTTCCCGCTGCTGCTGAGTGAAGGTTAAAGCCGGTATGAGACGCGCC 1928
```

```

Db      1396 ATGACAAAGATCCACCAATAGTAAGGCTCTGGCGGGGGGCGCATAGAGTG- GCGCAGAGACGA 1454
Qy      54  Asn-----GlyValThrIleVal-AsnGlyProLysGlySerGlyPheGlyGlyLeuAr 71
      ::: ::::
Db      1455 AGCTGTGGCGGTGTCCGTGCGTCTTAACGGT-----GCTTCGAGTTTGAAGGCTCTTCGA 1508
Qy      71  gserCyseGlyLysGlyTyrTrpLecysProCysCysAlaGlyLysValIleValAsnArgAl 91
      ::: ::::
Db      1509 AAATCTCACTCTCGCTGGGGGGGTCATCTCTGGCTGGAATTGAAGATCATTTGGGCGAAACCGCG 1568
Qy      91  aAspGlnIleSerGlnAlaValAlaHisGlnIleGlyThrGlySerValAlaMetValTh 111
      1569 CATTGAAGCTGGCTATTGCTCTTAAGAAATCACTTGGCGGGCGGGTGGCGCGCTCATGAAGATT 1628
Qy      111  rMet---ThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuThrPheGlyLeuSe 130
      ::: ::::
Db      1629 TGTGGGACACTGTTTGCACCAACCGCTCAAGATCATTTGGCGAGGTTGAAGCGGGTATTATA 1688
Qy      130  rAlaAlaTrpLysAlaAlaThrAsnGlyArgArgTyrArgThrGlnArgGlnMetYrG1 150
      1689 GACTGCGCTACTCTTCGATGGTGTGAAMAACATCTCAAGTGAAGAAAGAAACGTCGACCGTACGG 1748
Qy      150  yCyAspGlyTyrValArgAlaValGlnIleThrHisGly---LysAsnGlyTrpHisVal 169
      1749 GGTGGACACACCTATATGTGACTATAGAGTGCACAGACTCTTGGCGGAAACGGTGTGACATT 1808
Qy      169  IHisValHisAlaLeuLeuMetPheSerGlyAspValSerGlnAsnIleLeuGlnSerPh 189
      1809 GCACCGCAACATCGTGTGTTGTTCTTGATGTCGACCTGTCTGACATGAATCAAGGCATT 1868
Qy      189  eSerAspAlaMetPheAspArgTrpPheSerLysLeuValSerLeuGlyPheAlaAlaPr 209
      1869 TGAAGATTTCATGATTTTCCCGCTGGTCTGCTGATGTGTGAAGCCGATAGACGCGCC 1928
Qy      209  cLeuArgAsnSerGly---GlyLeuAsp---ValArgLysAlaGlyGlnIleAspG1 227
      1929 ACTGCGTAGACACGGGGTCAAACTTATCAAGATGTCTAACTCGGGGTGAAGCGCTGGCA 1988
Db      227  nValLeuAlaAlaTyrLeuThrTyrHisIleAlaSerGlyValGlyMetGlnValGlySerG1 247
      1989 A---ATGGCAACCTACTCCCTCAAG-----GGCATGTCTCAGAACTGATCGGCTC 2036
Qy      247  yAspGlyLysSerGlyArgHisGlyValAsnArgAlaProTrpGlnIleAlaValAspAlaVa 267
      2037 CGCTACTTAAACCGCGCTCTAAAGGGTGTGACACGCGCTTCCATGATG---TTGATATGTT 2092
Db      267  lGlyGlyAspProGlnIleLeuGln-----LeuTrpArgG1 279
      2094 GCGCGATCAAAAGCAAGCCCGCGAGAGATATAGACGCTGTTTGGTGCTCGGTGCGTGA 2153
Qy      279  uPheGlnPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeuArgAlaAr 299
      2154 GTATAGAGGTGTGTCTTAAAAACGTCGCTTG---TCTTGGTCACTGGG-----GCTAA 2204
Db      299  gAlaGlyLeuGlyValaGlnLeuThrAspAlaGlnIle-----ValGlnGlnGln-- 315
      2205 GCGTGTCTTGGGCACTTATATCATAGACGCTGATGTACGTCTGTAANAATGGAAGAAGAACT 2264
Qy      316  -----GlnSerAlaProValMetValAl 323
      2265 GTACAACATGCGCGGCTCTGGAAAGCAACCGGAACCGGGTGAATCAACCCGCGTGTCTTTC 2322
Db      323  aIleIleProAlaArgSerTrpMetMetIleArgThr 335
      2325 TTGTGTGAAGCCCGATGATTGGAACATGATTCAGTCT 2361

```

APPLICANT: ITO, HISAO  
 APPLICANT: GUNJI, YOSHIIYA  
 APPLICANT: YASUEBA, HISASHI  
 TITLE OF INVENTION: METHOD FOR PRODUCING L-ARGININE  
 FILE REFERENCE: 22531U50  
 CURRENT APPLICATION NUMBER: US/10/196,232  
 CURRENT FILING DATE: 2002-07-17  
 PRIOR APPLICATION NUMBER: JP 2001-224586  
 PRIOR FILING DATE: 2001-07-25  
 NUMBER OF SEQ ID NOS: 35  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 6  
 LENGTH: 4447  
 TYPE: DNA  
 ORGANISM: Brevibacterium lactofermentum  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1318)..(2598)  
 OTHER INFORMATION:  
 OS-10-196-232-6



QY 139 GYATGATGTTPATGThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaVal 158  
DB 1337 GCGCCGAGTGGCGGATTTACCAATTGTT-----GGGCACTCACATCAACT 1290  
QY 159 GGuileThrHisGly---LysAsnGlyTPrHISvalHisvalHisAlaLeu---LeuMet 176  
DB 1289 GAAATTAATCAATAGCTATGCAAAATGGGTGGACCCGCACTTTCACAAATTTGCTTGCTA 1230  
QY 177 PheSerGlyAspValSerGlu-----AsnIleLeuGluSerPheSerAspAlaMetPhe 194  
DB 1229 AAGAGTCGGCATGGAAGGCAAGCGCTGCATGATGCAAGCTTACGATGCG----- 1176  
QY 195 AspArgTPrHISerTyrSerLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGly 214  
DB 1175 -----TGCGTCGATGCTCTGAAGGGGTATGGCATGTGGGCG-----AACGATGCG 1131  
QY 215 GYLeuAspValArgGlySerIleGlyGlyAlaAspGlnValLeuAlaIleThr 234  
DB 1130 GCGTTTGAAGTTCGCGGGGCTGCCAATGCGGCGAC-----TACGTAGCG 1086  
QY 235 LysIleAlaSerGlyValGlyMetGluValGlySerGlyAspGlyLysSerGlyArgHis 254  
DB 1085 AAG-----TGGGGCGCTCCGCAAGAGTTCAGCTTTCAGCTCCGAAGCGGAAACCG 1032  
QY 255 GYAsnArgAlaProTPrGluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeu 274  
DB 1031 AAGAGCAGCAGCCAGGAGGAGCTACT-----CAGCGCGGCGCATAC----- 990  
QY 275 GGuileTPrArgGluPheGluPheGlySerMetGly-----ArgArgAlaIleAlaTPr 292  
DB 989 GGGCTTTGGCTCGCAATATTTCATGCAAGCGGCGGAGCTGAGCGCAATTGGTGTGG 930  
QY 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGlyLeuThrAsp----- 308  
DB 929 AGCCAAGGCTTGAAGAAAGAAATCCGCTTG---GTGAGCTGATATCAATGAAGCTATG 873  
QY 309 AlaGlnIleValGluGlnGluGlnSerAlaProValMetValAlaIleIleProAlaArg 328  
DB 872 GCTGAAGTTGATGACAGACAGGCAAGCGGCTTGAATATTGCGGAGTGGGCAATGAA 813  
QY 329 SerTPrMetMetIleArg 334  
DB 812 GGGTGAAGCAAGTGGCG 795  
RESULT 14  
US-10-263-666-6/c  
; Sequence 6, Application US/10263666  
; Publication No. US20030073224A1  
; GENERAL INFORMATION:  
; APPLICANT: D'Elia, John  
; APPLICANT: Stoddard, Steven F.  
; TITLE OF INVENTION: KETOGLUCONIC ACID ENDGENOUS PLASMTDS  
; FILE REFERENCE: 1533.1260004  
; CURRENT APPLICATION NUMBER: US/10/263,666  
; PRIOR FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: US 09/826,191  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: US 60/194,627  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent version 3.0  
; SEQ ID NO 6  
; LENGTH: 2401  
; TYPE: DNA  
; ORGANISM: Ketogluconigenium  
US-10-263-666-6  
Alignment Scores:  
Pred. No.: 2,15e-21 Length: 2401  
Score: 282.00 Matches: 86  
Percent Similarity: 44.41% Conservative: 41  
Best Local Similarity: 30.07% Mismatches: 123  
Query Match: 14.06% Indels: 36

DB: 14 Gaps: 14  
US-10-007-527a-2 (1-379) x US-10-263-666-6 (1-2401)  
QY 60 AsnGlyProLysGlySerGlyPheGlyLysLeuArgSerCysGlyLysGlyTPrIleCys 79  
DB 1577 GATGGCGGGCGCTGGGCGCTTCTTTCAGCTTATGACAGAGTCCGGAATGTGGGGCTGC 1518  
QY 80 ProCysCysAlaGlyLysValGlyValHisArgAlaAspGluIleSerGlnValAla 99  
DB 1517 GCAATTTGAGCGCTGCATGCGCGCAATTCGCGCTTGAATGAAATCACTGCTCGCG 1458  
QY 100 HisGlnLeuGlyTPrGlySerVal---AlaMetValIleThrMetHisThrAla 118  
DB 1457 TGGGGCGCGCAAGAGGCTTGTGTGCGGTCTCAAACTGTACAGCCGCAAGCAAGCA 1398  
QY 119 GYLeuArgLeuHisAspLeuTPrHISerLeuSerAlaIleTPrLysAlaIleThrAsn 138  
DB 1397 GGTGATGGCTTTTGAATCTTCTCCAAACATGAAGAGGCAAGCGGCTTGGCTCG 1338  
QY 139 GYATGATGTPrArgTPrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaVal 158  
DB 1337 GCGCGAGTGGCGGATTTACCAATTGTT-----GGGCACTCACATCAACT 1290  
QY 159 GGuileThrHisGly---LysAsnGlyTPrHISvalHisvalHisAlaLeu---LeuMet 176  
DB 1289 GAAATTAATCAATAGCTATGCAAAATGGGTGGACCCGCACTTTCACAAATTTGCTTGCTA 1230  
QY 177 PheSerGlyAspValSerGlu-----AsnIleLeuGluSerPheSerAspAlaMetPhe 194  
DB 1229 AAGAGTCGGCATGGAAGGCAAGCGCTGCATGATGCAAGCTTACGATGCG----- 1176  
QY 195 AspArgTPrHISerTyrSerLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGly 214  
DB 1175 -----TGCGTCGATGCTCTGAAGGGGTATGGCATGTGGGCG-----AACGATGCG 1131  
QY 215 GYLeuAspValArgGlySerIleGlyGlyAlaAspGlnValLeuAlaIleThr 234  
DB 1130 GCGTTTGAAGTTCGCGGGGCTGCCAATGCGGCGAC-----TACGTAGCG 1086  
QY 235 LysIleAlaSerGlyValGlyMetGluValGlySerGlyAspGlyLysSerGlyArgHis 254  
DB 1085 AAG-----TGGGGCGCTCCGCAAGAGTTCAGCTTTCAGCTCCGAAGCGGAAACCG 1032  
QY 255 GYAsnArgAlaProTPrGluIleAlaValAspAlaValGlyGlyAspProGlnAlaLeu 274  
DB 1031 AAGAGCAGCAGCCAGGAGGAGCTACT-----CAGCGCGGCGCATAC----- 990  
QY 275 GGuileTPrArgGluPheGluPheGlySerMetGly-----ArgArgAlaIleAlaTPr 292  
DB 989 GGGCTTTGGCTCGCAATATTTCATGCAAGCGGCGGAGCTGAGCGCAATTGGTGTGG 930  
QY 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGlyLeuThrAsp----- 308  
DB 929 AGCCAAGGCTTGAAGAAAGAAATCCGCTTG---GTGAGCTGATATCAATGAAGCTATG 873  
QY 309 AlaGlnIleValGluGlnGluGlnSerAlaProValMetValAlaIleIleProAlaArg 328  
DB 872 GCTGAAGTTGATGACAGACAGGCAAGCGGCTTGAATATTGCGGAGTGGGCAATGAA 813  
QY 329 SerTPrMetMetIleArg 334  
DB 812 GGGTGAAGCAAGTGGCG 795  
RESULT 15  
US-10-261-481-6/c  
; Sequence 6, Application US/10261481  
; Publication No. US20030077830A1  
; GENERAL INFORMATION:  
; APPLICANT: D'Elia, John  
; APPLICANT: Stoddard, Steven F.  
; TITLE OF INVENTION: KETOGLUCONIC ACID ENDGENOUS PLASMTDS  
; FILE REFERENCE: 1533.1260003



; CURRENT APPLICATION NUMBER: US/10/261,481  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: US 09/826,191  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: US 60/194,627  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 6  
; LENGTH: 2401  
; TYPE: DNA  
; ORGANISM: Ketogulonigenium  
US-10-261-481-6

## Alignment Scores:

Pred. No.:	2,15e-21	Length:	2401
Score:	282.00	Matches:	86
Percent Similarity:	44.41%	Conservative:	41
Best Local Similarity:	30.07%	Mismatches:	123
Query Match:	14.06%	Indels:	36
DB:	14	Gaps:	14

US-10-007-527a-2 (1-379) x US-10-261-481-6 (1-2401)

QY 60 AsnGlyProLysGlySerGlyPheGlyGlyLeuArgSerGlyLysGlyTyrPheCys 79  
DB 1577 GATGGCGGCGGTCGCGCATCTTTTCAGCTAGCAGTCGCGAATGTGTGGGCTGC 1518  
QY 80 ProCysCyAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAla 99  
DB 1517 GCAGTTGTAGCGCTGCGATCGCGCAAAATTCGCGTTTGAATAATCATCTTCTCGCG 1458  
QY 100 HisGlnLeuGlyThrGlySerVal--AlaMetValThrMetThrMetArgHisThrAla 118  
DB 1457 TGGGCGCGCGACACGCGTTTGTCCGCGCTCATTACTGTGACAGCCAGCAACAAGCA 1398  
QY 119 GlyGlnArgLeuHisAspLeuTyrPheGlyLeuSerAlaAlaTyrLysAlaAlaThrAsn 138  
DB 1397 GGTGATTCGCTTTTATGATCTTCTCCAAACATGAGAGGCGAAACAGCGTTGCGTCAG 1338  
QY 139 GlyArgArgTyrPheArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaVal 158  
DB 1337 CGCCGCGAGTGGCGCATTTTACATTTTGT--GAGTCAATCATCATCACT 1290  
QY 159 GluIleThrHisGly--LysAsnGlyTyrHisValHisValHisAlaLeu--LeuMet 176  
DB 1289 GAAATTACTCATAGCTATGCAAAATGGGTGGCAACCGCACTTTCAGAAATGTCTGTGCTA 1230  
QY 177 PheSerGlyAspValSerGlu-----AsnIleLeuGluSerPheSerAspAlaMetPhe 194  
DB 1229 AGAGCTGGCGATGAAAGCGAGCGCTGATGATGCAACGTCTAGGCGATGCG----- 1176  
QY 195 AspArgTyrThrSerLysLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGly 214  
DB 1175 -----TGGCGTGCATGTCTGAAGGGGTATGGCATGTGGGCG-----AACGATGCG 1131  
QY 215 GlyLeuAspValArgLysIleGlyGlyGluAlaAspGlnValIleuAlaAlaTyrLeuThr 234  
DB 1130 GCGTTTACCTTGGCGGCGCTGCCAATGCCGCGGAC-----TACGTAGCG 1086  
QY 235 LysIleAlaSerGlyValGlyMetGluValGlySerGlyAspGlyLysSerGlyArgHis 254  
DB 1085 AAG-----TGGGCGCTGCCGAAAGAGTTGACGCTTCAAGCTCGAAGACGGGAAACGC 1032  
QY 255 GlyAsnArgAlaProTyrGluIleAlaValAspAlaValGlyLysAspProGlnAlaLeu 274  
DB 1031 AAAGAGCGACGCCAAGCGACGTACTT-----CAGCGGCGCATGAC----- 990  
QY 275 GluLeuTyrArgGluPheGluPheGlySerMetGly-----ArgArgAlaIleAlaTyr 292  
DB 989 GGGCTTGGCTCGAATTTTCATGCAACAGCGCAAGCGTACGCGCATTTGTGTGG 930  
QY 293 SerArgGlyLeuArgAlaArgAlaGlyLeuGlyValGluLeuThrAsp----- 308

DB 929 ACCCAAGCTTGAAGAAAGAAATCGCGTTG--GTGAGCTGTGATGACGATGAACCTATG 873  
QY 309 AlaGlnIleValGluGlnGluSerLysProValMetValAlaIleIleProAlaArg 328  
DB 872 GGTGAAGTGTGATGCGACAGAGCAAGCGCCCTGAAATTTGTCCGCGAGTGGGACATGAA 813  
QY 329 SerTyrMetMetIleArg 334  
DB 812 GGTGGAAGCAAGTGC 795

Search completed: August 22, 2003, 12:13:47  
Job time : 267 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 22, 2003, 10:02:44 ; Search time 1920 Seconds  
(without alignments)  
4797.602 Million cell updates/sec

Title: US-10-007-527a-2  
Perfect score: 2005  
Sequence: 1 MTSVSAEHLGSKDRPPVLVS.....HLYRIIPADVPPIISVAK 379

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Rgapop 6.0 , Rgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgm2\_1/USPTO.spool/US10007527/runat\_20082003\_165152\_15377/app\_query.fasta\_1.519  
-DB=BSR -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdd -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEADSIDE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10007527 @CGN 1.1 2135 @runat\_20082003\_165152\_15377 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUTPUT -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DIV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_hic3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	129.5	6.5	929	13	BUS19534	BUS19534 AGENCOURT
C 2	129	6.4	1648	12	BM913094	BM913094 AGENCOURT
C 3	128.5	6.4	1090	12	BM457396	BM457396 AGENCOURT
C 4	122.5	6.1	894	12	BI661991	BI661991 AGENCOURT
C 5	119.5	6.0	1661	29	CC321442	CC321442 TAM32-34D
C 6	119	5.9	950	13	BU271357	BU271357 603818506
C 7	119	5.9	1430	12	BM907785	BM907785 AGENCOURT
C 8	119	5.9	1687	12	BI224374	BI224374 602940916
C 9	118	5.9	851	13	BI714786	BI714786 AGENCOURT
C 10	118	5.9	893	29	BE554247	BE554247 pacsl-60
C 11	117	5.8	1306	13	BQ936859	BQ936859 AGENCOURT
C 12	117	5.8	1615	10	BF783863	BF783863 602109787
C 13	116.5	5.8	689	13	BU705875	BU705875 UI-M-FRO-
C 14	116	5.8	867	12	BI411296	BI411296 602964678
C 15	116	5.8	1399	10	BS684181	BS684181 602635811
C 16	115.5	5.8	703	14	CS246284	CS246284 UI-M-FD-
C 17	115.5	5.8	786	13	BQ572893	BQ572893 UI-M-FD-
C 18	115.5	5.8	799	14	CA320693	CA320693 UI-M-FW-
C 19	115.5	5.8	857	14	CA750314	CA750314 UI-M-FD-
C 20	115.5	5.8	918	14	BY717387	BY717387 BY717387
C 21	115.5	5.8	2161	11	AK049163	AK049163 Mus muscu
C 22	115.5	5.8	2161	11	AK083015	AK083015 Mus muscu
C 23	115.5	5.8	2162	11	AK028287	AK028287 Mus muscu
C 24	115.5	5.8	2167	11	BC013842	BC013842 Mus muscu
C 25	115.5	5.8	2168	11	AK017609	AK017609 Mus muscu
C 26	115	5.7	696	14	BY729572	BY729572 BY729572
C 27	115	5.7	899	29	BZ551248	BZ551248 pacsl-60
C 28	114.5	5.7	930	12	BS922360	BS922360 602820270
C 29	114.5	5.7	1189	29	CC318464	CC318464 TAM32-32C
C 30	113.5	5.7	974	13	BO672428	BO672428 AGENCOURT
C 31	113.5	5.7	1414	13	BO267920	BO267920 AGENCOURT
C 32	113	5.6	932	10	BG444776	BG444776 CA_Ba002
C 33	113	5.6	1186	10	BF786408	BF786408 602110681
C 34	112.5	5.6	1279	10	BF160413	BF160413 601771374
C 35	112	5.6	1480	13	BU434837	BU434837 603220135
C 36	111.5	5.6	926	12	BM457074	BM457074 AGENCOURT
C 37	111.5	5.5	1221	29	CC297234	CC297234 CH261-134
C 38	110.5	5.5	984	12	BI655429	BI655429 603285076
C 39	110.5	5.5	1242	10	BE568031	BE568031 601341540
C 40	110.5	5.5	1528	12	BM904431	BM904431 AGENCOURT
C 41	110	5.5	1254	29	CC221328	CC221328 CH261-150
C 42	110	5.5	1297	29	AG111098	AG111098 Pan trogl
C 43	109.5	5.5	1157	13	BO960888	BO960888 AGENCOURT
C 44	109.5	5.5	1682	12	BM457385	BM457385 AGENCOURT
C 45	109	5.4	874	29	CC436734	CC436734 PUDCP391TD

#### ALIGNMENTS

RESULT 1  
BUS19534/c 929 bp mRNA linear EST 12-SEP-2002  
LOCUS BUS19534  
DEFINITION AGENCOURT 10166703 NIH MGC 134 Mus musculus cDNA clone  
IMAGE:6517370 5', mRNA sequence.  
ACCESSION BUS19534  
VERSION BUS19534.1 GI:22827060  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 929)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>,  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Dr. David Rowe  
 cDNA Library Preparation: Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L14M14097 row: g column: 03  
 High quality sequence start: 6  
 High quality sequence stop: 710.  
 Location/Qualifiers

## FEATURES

source 1.929  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6517370"  
 /tissue\_type="undifferentiated limb"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NIH\_MGC\_134"  
 /note="Vector: pCMV-SPORT6.1.cdb; Site\_1: EcoRV; Site\_2:  
 NotI; Cloned unidirectionally. Primer: Oligo dT. Average  
 insert size 1.7 kb. Constructed by Resgen, Invitrogen  
 Corp. Note: this is a NIH MGC library."  
 BASE COUNT 183 a 312 c 251 g 183 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 0.0953 Length: 929  
 Score: 129.50 Matches: 91  
 Percent Similarity: 31.90% Conservative: 28  
 Best Local Similarity: 24.40% Mismatches: 95  
 Query Match: 6.46% Indels: 160  
 DB: 13 Gaps: 19

US-10-007-527a-2 (1-379) x BUS19534 (1-929)

QY 20 SerSerAspLyLeaArgGlyLeaArgHisGluLeuArgProlys-----leu 34  
 DB 768 GCTTCTTCAAAAAGCCGCTTCGCGCATCATCTT---CTGACACCCCTTGCTTGACTTGG 712  
 QY 35 GlnGlnIleThrSerGluThrPrope-AspAlaCysGlyArgProIleSerGlyValAs 54  
 DB 711 CAAAGAGTCACAGTCTCAGAGCATGTGTGAGCATGCTCACA----- 669  
 QY 54 nglyValThrIleValAsnGlyProlysGlySerGlyPheGlyGlyLeuArgSerCysG 74  
 DB 669 ----- 669  
 QY 74 yLysGlyTyrPrlCysProCysCysAlaGlyLysValGlyAlaHisArgAlaAspGlu 94  
 DB 668 -----TCG-----TGTCCTAGCTGTCG----- 654  
 QY 94 eSerGlnValValAlaHisGlnLeuGlyThrGlySerValAlaMetValThrMetThrme 114  
 DB 653 -----GGAACATCTCAACTGACACCTGATCTGTAGAGGT 619  
 QY 114 tArgHisThrAlaGlyGlnArgLeuHisAspLeuTyrThrGlyLeuSerAlaAlaTrrply 134  
 DB 618 A-----GGAACAGGCGCGGAGACCA-----GGTCGAGCCAGAGCATGGA 578  
 QY 134 sAlaAlaThrAsnGlyArgArgTyrArgThrGlu-----ArgGluMetTyrGly-Cys 151  
 DB 577 G-----CGCTGGTAGGACAGAGCTTCTCCACACATGATGTTGCTCTGT 536  
 QY 152 AspGlyTyrValArgAlaValGluIleThrHisGlyLysAsnGly---TrpHisValHis 170  
 DB 535 -----GGGCACTTGACATCTGGGGTACTAGG 509

QY 171 ValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPheSer 190  
 DB 508 CTCACAGCATTCCTTGACTCGAGC----- 482  
 QY 191 AspAlaMetPheAspArgTyrThrSerTyrValSerLeuGlyPheAlaAlaProleu 210  
 DB 481 -----TCTCAGAGGCCCGTGGCTGCCACTTCCGCGC 446  
 QY 211 ArgAsnSerGlyGlyLeuAspValArgIleValIleGlyGlyAlaAspGlnValleuAla 230  
 DB 445 CCGCTTCAGGACGACTGAGCTCAGA-----GACGAGCTCTCTGA 404  
 QY 231 AlaTyrLeuThrIleAlaSerGlyVal---GlyMetGluValGlySerGlyAspGly 249  
 DB 403 GGGATCTTGCCCGGCTGTGACCAACATTAACAGTTGACTCTTGTTCGCTTGACGAG 344  
 QY 250 LysSerGlyArgHisGlyAsnArgAlaProTyrGluIleAlaValAspAlaValGlyGly 269  
 DB 343 GAGTAGGCAAAAGGGGAGCCAGAGGTCAAGCTGAGAGGCTAGAGGTACG----- 296  
 QY 270 AspProGlnAlaLeuGluLeuTyrArgGluPheGluPheGlySerMetGlyArgArgAla 289  
 DB 295 -----TCG-----AGGCTGGG 284  
 QY 290 IleAlaTyrSerArgGlyLeuArgAlaArgAlaGlyLeuGlyAlaGlyLeuThrAspAla 309  
 DB 283 CTCACCTGGGAACTGGGCTCAGGCAAGAACTGGGCGCAGG----- 242  
 QY 310 GlnIleValGlnGluSerAlaProValMetValAlaIleIleProAlaArgSer 329  
 DB 241 -----GTCCATCCAGGACGCCCTTACT-----CCGGCGCGC 209  
 QY 330 TrpMetMetIleArgThrCysAlaProTyrValPheGlyGlnIleLeuGlyLeuValGlu 349  
 DB 208 -----GAGCGGCGCGGCTGTGACT 188  
 QY 350 AlaGlyAlaThrTyr-GluAsnLeuArgAspHisIleuHisTyrArgGluProAlaAlaAs 369  
 DB 187 CCGGAGTTCGCTGGCGGCGCGGCGCTTGGCGCTGCTCCGGGTGAACGACGAGAC 128  
 QY 369 pValArgProProlIleIleSerValArgLys 379  
 DB 127 CAGAGGCCACAGGCGCTGGGCTGAGACAG 97

RESULT 2  
 BM913094/c 1648 bp mRNA linear EST 12-MAR-2002  
 LOCUS AGENCORRT 6613820 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5475102  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BM913094  
 VERSION BM913094.1 GI:19363473  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1648)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: DCTD/DRP  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L14M1989 row: 5 column: 07  
 High quality sequence stop: 263.  
 Location/Qualifiers  
 source 1.1648

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5475102"
/issue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH_MGC_41"
/note="Organ: Skin; Vector: pOTB7; Site 1: XhoI, Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."
BASE COUNT      333 a      668 c      437 g      205 t      5 others
ORIGIN
```

## Alignment Scores:

```
Pred. No.:      0.227      Length:      1648
Score:          129.00     Matches:      93
Percent Similarity: 30.73%   Conservative: 33
Best Local Similarity: 22.68% Mismatches:     153
Query Match:      5.43%     Indels:       131
DB:              12        Gaps:         17
```

US-10-007-527a-2 (1-379) x BM913094 (1-1648)

```
QY      10 SerGlyAspArgProPheValLeuValSerSerAspIysArgIleArgHisGlu 29
      1101 GCTGGCCCAAGACGGCCGCCAATCAATATATACC-----CCACTGGGGCTCCGACACCGA 1048
QY      30 LeuArgProLys-----
      1047 AAGCGTCCCAAGCGGTGGGCGACAGAGAGCGGTGACTGCTGCTGCTTATGACAGATTGG 988
QY      34 ---LeuGlnGlnIleThrThrSerGluThrPheAsnAlaCysGlyArgPro----- 49
      987 CCGTGGAGCAACTCTCCACAGNAG-----GGGCGCGGTGAGACTGTA 946
QY      50 -----IleSerGlyValAsn-----
      945 AGTTGCTAGAGGGGCGGTATATCTGCGCTGCGCGGATGATAGTGACCGACAGAGAGAG 886
QY      55 -----
      885 GTGTGATTGTGCTCGGTGCGCGCTTAGTAGGCACAGAGTGTGTGGACACATGGT 826
QY      65 SerGlyPheGlyGlyLeu-ArgSerCys-----GlyLysGlyT 77
      825 GCGGGTCTGTCTGGGCTCAGTGTGCTGGGGCTGGGAGTCCCGTCTGGGCGGTGGAT 766
QY      77 TrpLecCysProCys-----CysAlaGlyLysValGlyAlaHisArg 91
      765 GGGGTTGTGGTTGGCGGCATTGGAGCGCGCTGTGTGCGCGCATCTTGGTGTGGCGCG 706
QY      91 LaaAspIuIleSerGlnValAlaHisGlnLeuGly--ThrGlySerValAlaMet 110
      705 GATCTGTTTGGCTGCTGCGCGCACTAGCATGCGGCGCGCGGTGTGTCATGGGCA 646
QY      110 aThrMetThrMetAlaGlnHisThrAlaGlyGlnArgLeuHisAspLeuTrpThrGlyLeu 130
      645 TTTCAGCCGAGCGCGCGCATCCCGCC-----GAGGGGCGCG 610
QY      130 eraIaIaIaTrpLysAlaIaIaThrAsnGlyValArgArgTrpArgThrGlnArgGluMet 150
      609 GGGGGAATCTGG---GTTGCGCGCGCGACAGCCCATGTGGCGCTTGGCGCGCAGCTGG 553
QY      150 IyCysAspGlyTyrValArgAlaValGluIleThrHisGlyLysAsnGly-TrpHisVal 169
      552 GGGGCGGTGGGC-----GTTAGTGGCGGTTTGGCACTCC 520
QY      170 HisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGlnUserPhe 189
```

```
DB      519 GAGACGGGGGGGATGCTGTGACTGTGG-----
QY      190 SerAspAlaMetPheAspArgTrpThrSerLysLeu-----
DB      489 -----TGAACGGGCGTACAGCGAGCTTCGTGGCGGCTCCCGGT 451
QY      202 ---ValSerLeuGly-PheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArg 220
      450 GGTGTGCGCGTTGGCGCTTGAACCGCGCGCGCGTATAGTGGACAGCGTCCGCTTTC 391
QY      220 sIleGlyGlyGlnAlaAspGlnValLeuAlaIaIaTrpLeuThrLysIleAlaSerGly 240
      390 TCGTGGCGGTGCGCTTGAAGTAAAGTTCGGGGGCGTGTGCACACATTTTGTGC----- 336
QY      240 IGIyMetGluValGlySerGlyAspIyAspIyLysSerGlyArgHisGly-----As 256
      335 -GACATGTGTGCGCGCGCGCGCGCGCGCTGTGTGGGAGCCGGGTCCCGGACCGAGAC 277
QY      256 nArgAlaProTrpGluIleAlaValAspAlaValGlyIAspProGlnAlaLeuGlu 276
      276 GAGGGGCGCTGTG-----ACGGCGAGAGGTGGCCGCGCGGAACTTGCGCG 229
QY      276 uTrpArgGluPheGlu-PheGlySerMetGlyArgArgAlaIleAlaTrp----- 292
      228 GTGGCGCGCGCGCGCGCGACAGCGCGAGCGCGCGCTGTGCAACTGTGCTCTGGA 169
QY      293 ---SerArgIyLeuArgAlaArgAlaGlyLeuGlyValGluLeuThrAspAlaGln 311
      168 GAGCGCGCGAGAGCGCGCGCGGTGGCGCGCGGTGTGTGAGCGGAGACCGCGAGCGG 109
QY      311 IeValGlnGlnGlnGlnSerAlaProValMetValAlaIleIleProAlaArgSerTrp 331
      108 TCGGCGCTGTCCGAGAGCGCGCGCGGAAAGCGCGGTGACAGCGGTGCGCGCGCTCGGCG 49
QY      331 eMetIleArgThrCysAla 337
DB      48 CCGTAGCTCATCTCCGGGCT 29
```

```
RESULT 3
BM457396      1090 bp      mRNA      linear      EST 05-FEB-2002
LOCUS
DEFINITION
AGENCOURT 6407558 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5584861
5', mRNA sequence.
ACCESSION
BM457396
VERSION
BM457396.1 GI:18506436
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1090)
NIH-MGC http://img.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM12350 row: d column: 14
High quality sequence drop: 538.
Location/Qualifiers
```

## FEATURES

source

```
1..1090
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5584861"
/tissue_type="embryonal carcinoma, cell line"
```



DB: 12 Gaps: 16

US-10-007-527a-2 (1-379) x B161991 (1-894)

QY 68 GlyGlyLeuArgSerCysGlyGlyGly-----Trr-----IleCysPro 80  
 DB 811 GGGACATTGACCTCTCGCGGAGACGATACAGCTGGGATACAGGTTTTCATTTCCCA 752  
 QY 81 CysCysAla-----GlyLysValGlyAlaHisArgAlaAspGluIleSer 95  
 DB 751 CATTGGCTCTTCAAGACGCTTGCAGATCATCTCTCGACACCTTGAC----- 701  
 QY 96 GlnValValAlaHisGlnLeuGlyThrGlySerValAlaMetValThrMetThrMetArg 115  
 DB 700 -----TTTGGCAAGGTCACAGCTCAGACGATTTGGGAGCATGCT 659  
 QY 116 HisThr-----AlaGlyGlnArgLeu-----HisAspLeuTrpThrGlyLeuSer 130  
 DB 658 CACAAAGGTGTCCATGCTGCGGACATCTCACTACATGACCTGTGATGAGGTAGAC 599  
 QY 131 AlaAlaTrpLysAla-----AlaThrAsnGlyArgArgTrpArg 143  
 DB 598 AAGGCTGGGAGACGAGCTGAGCAGACAGATGAAAGCGCTGTAGGACGAGCTTTC 539  
 QY 144 ThrGluArgGluMetArgGlyCysAspGlyTrpValArgAlaValGluIleThrHisGly 163  
 DB 538 GACACATGTTGCTGTGGGACCTTGGCATC----- 506  
 QY 164 LysAsnGlyTrpHisValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGlu 183  
 DB 505 -----TGGGTACTAGGCTCCACAGATCTTGTGACTGGGC----- 467  
 QY 184 AsnIleLeuGluSerPheSerAspAlaMetPheAspArgTrpThrSerLysLeuValSer 203  
 DB 466 -----TCTCAGGGCCCGTGC 452  
 QY 204 LeuGlyPheAlaAlaProLeuArgAsnSerGlyGlyLeuAspValArgLysIleGlyGly 223  
 DB 451 CTGGGCTCCCACTTCCGTCGCCCTCAGCAGCAGCTGAGCTCAGA----- 404  
 QY 224 GluAlaAspGlnValLeuAlaAlaValTrpLeuThrLysIleAlaSerGlyVal---GlyMet 242  
 DB 403 -----GACGAGTCTCTGAGGGGATCTTGGCGCTGTGCAACCACTAAACAGTTGT 350  
 QY 243 GluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTrpGluIle 262  
 DB 349 ACTCTTGTTCGCTTGAACGGGAGATGGGCAACGGGAGGACCGGCTGAGGGGCT 290  
 QY 263 AlaValAspAlaValGlyGlyAspProGlnAlaLeuGluLeuTrpArgGluPheGluPhe 282  
 DB 289 AGGCTCAGC-----TGG----- 278  
 QY 283 GlySerMetClyArgArgAlaIleAlaIleAlaTrpSerArgGlyLeuArgAlaArgAlaGlyLeu 302  
 DB 277 -----AGGCTGGGCTCAGCTGGGGAACCTGGCTCAGCAGAGAACTGGGGCA 230  
 QY 303 GlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluSerAlaProValMetVal 322  
 DB 229 GGA-----GTCCATCCAGGACGCGTAACT----- 203  
 QY 323 AlaIleIleProAlaArgSerTrpMetMetIleArgThrCysAlaProTrpValPheGly 342  
 DB 202 -----CCGGCGCGC----- 194  
 QY 343 GlnIleLeuGlyLeuValGluAlaGlyAlaThrTrp-GluAsnLeuArgAspHisIleuHis 362  
 DB 193 GACGGCGGGGTGGTGCATCTCCGGGCTTCGGCTGGGGCGCGGCTTGGCGGCTGCT 134  
 QY 362 sTrArgLeuProAlaAlaAspValArgProPheIleSerValArgLys 379  
 DB 133 GCTCCGGGTGAACGACGAGCCACGAGGCCACAGGGCTGGGGGTGAAGACG 82

RESULT 5

CC321442/c

LOCUS CC321442 1661 bp DNA linear GSS 14-MAY-2003

DEFINITION TAM32-34D5\_ECI.1 TAM32 Gallus gallus genomic clone TAM32-34D5,

ACCESSION genomic survey sequence.

CC321442

VERSION CC321442.1 GI:30715500

KEYWORDS GSS.

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

1 (bases 1 to 1661)

REFERENCE 1 Kremetzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.

TITLE Gallus gallus BAC End Reads

JOURNAL Unpublished

COMMENT Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@watson.wustl.edu

Insert Length: 15200 Std Error: 0.00

Seq primer: ECI TACGACTCATATAGGGCG

Class: BAC ends

High quality sequence start: 74

High quality sequence stop: 124.

Location/Qualifiers

1..1661

/organism="gallus gallus"

/mol\_type="genomic DNA"

/strain="Red Jungle Fowl"

/db\_xref="taxon:9031"

/clone="TAM32-34D5"

/sex="female"

/cell\_line="WUCD001, Inbred 256"

/clone\_1ib="TAM32"

/note="Vector: pBCEBAC1, Site 1: EcoRI, Site 2: EcoRI; TAM32 Female Chicken Library - for library and clone ordering information: http://www.hbz.tamu.edu"

BASE COUNT 311 a 949 c 157 g 238 t 6 others

ORIGIN

Alignment Scores:

Pred. No.: 1.71 Length: 1661

Score: 119.50 Matches: 77

Percent Similarity: 33.23 Conservative: 31

Best local Similarity: 23.69 Mismatches: 118

Query Match: 5.968 Indels: 99

DB: 29 Gaps: 14

US-10-007-527a-2 (1-379) x CC321442 (1-1661)

QY 54 AsnGlyValThrIleValAsnGlyProLysGlySerGlyPheGlyGlyLeuArg----- 71  
 DB 1492 GATGCGTGAACGAGTCCGTGTGAGAGGGGAGTGAATCAGAGGCCAGAGGGTGTTC 1433  
 QY 72 -----Ser 72  
 DB 1432 ATGTGGAGAGGAGCGAGAGCGAGGCTGGCTTGTGAGAGGAGATATATGAGGAGGT 1373  
 QY 73 CysGlyLysGlyTrpIleCysProCysCysValaGlyLysValGlyAlaHisArgAlaAsp 92  
 DB 1372 GTGGGAGAGGTAAATGAGTGT-----GTGAGTGGCGGGAT 1337  
 QY 93 GluIleSerClnValValAlaHisGlnLeuGlyThrGlySerValAlaMetValThrMet 112  
 DB 1336 ACGAGAGGTGTTGTGTGCATGATCGTGGGCGAGGGGGCGGGTCCGAGTGGGCTCG 1277  
 QY 113 ThrMetArgHisThrAlaGlyGlnArgLeu-HisAspLeuTrpThrGly----- 128  
 DB 1276 GGGGCGTGAAGTGTGAGGGGGCAACGGGTAGAGG-----TGGGAGGTGGGGGTGGAG 1223  
 QY 129 ----LeuSerValAlaIleTrpLys---AlaAlaThrAsnGlyArgArgTrpArgThrGluArg 146





Db 562 -----TGTGTGGGGCTGG 550  
QY 207 laAlaProLeuAaGaaSereGlyLeuAspValArgLysIleGlyValAlaAspG 227  
Db 549 TTTCAGAGCTGCGCTGGGTTGGTGGCTCCGGGTTTGGCTGGGTTGGCTGGG 490  
QY 227 InVal-----LeuAlaAlaTrpLeuThrLysIleAlaSerGlyValG 241  
Db 489 CCGTGAAGTTCAGTGGCTGGGGCCATCTCAGTGAAGCTTGAAGCTTGGGATCAG 430  
QY 241 lMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTArg 261  
Db 429 GCGTGGAGCAGGGGTTGGA-----TGGC 406  
QY 261 lulleAlaValAspAlaValAlGlyLysProGlnAlaLeuGluLeuTArgGluPheG 281  
Db 405 GAGTGCAGGTAGAGAGCAAGTCTGGCAA----- 377  
QY 281 lPheGlySerMetGlyArgArgAlaAlaLeuT-----SerArgGlyLeuArgAla 299  
Db 376 -----GGCCCCGAGGCAAGGCCGTTAGGGCTTGTATGCTTGCATGGGATCGCAG 322  
QY 299 rGlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGluSerAlaP 319  
Db 321 GAGCA-----TCAGCCG 310  
QY 319 roValMetValAlaIleIleProAlaArg-----SerTrpMetMetIleA 334  
Db 309 CGGGGCTCAGAGTTGCTTCCACACGAGCTTCAGATGAAGCTTCATGAGCTTCAGG 250  
QY 334 rGhrCysAlaProTyrValPheGlyGluLeuGlyLeuValGlu----- 349  
Db 249 CAAGAAGCTGAGTTGCACTTGCAGGCAAGTCTGTGGGCGCCAGCAACACAGAGA 190  
QY 350 --AlaGlyAlaThrTrpGluAsnLeuArgAspHisLeuHisTyrArg-----LeuP 366  
Db 189 GGGCTGGACCAACAGATGATGATCTTTCAGCAGCATTTAAAGCCCTGCTCTCC 130  
QY 366 roAlaAlaAspValArgProPro 373  
Db 129 CTGACAGCATG-GAGCGTCCGCT 108  
RESULT 7  
BM907785/c 1430 bp mRNA linear EST 12-MAR-2002  
LOCUS BM907785  
DEFINITION AGENCOURT 6618957 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:5736201  
5', mRNA sequence.  
ACCESSION BM907785  
VERSION BM907785.1 GI:19358164  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1430)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs@email.nih.gov](mailto:cgabs@email.nih.gov)  
Tissue Procurement: DCTD/DTF/Gasdar  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LIML at:  
<http://image.llnl.gov>  
Plate: LIML2744 row: f column: 10  
High quality sequence stop: 319.  
Location/Qualifiers  
1..1430  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db xref="taxon:9606"  
/clone="IMAGE:5736201"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 68"  
/note="Organ: lung; Vector: pCMV-SPORE; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."

BASE COUNT 341 a 598 c 228 g 258 t 5 others  
ORIGIN

Alignment Scores:  
Pred. No.: 1.56 Length: 1430  
Score: 119.00 Matches: 90  
Percent Similarity: 34.82% Conservative: 43  
Best Local Similarity: 23.56% Mismatches: 110  
Query Match: 5.94% Indels: 139  
DB: 12 Gaps: 21

US-10-007-527A-2 (1-379) x BM907785 (1-1430)

QY 46 CysGlyArgProIleSerGlyValAsnGlyValThrIleValAsnGlyProLysGlySer 65  
Db 1343 TGTGGCGCCCA-----GATGGGGGGGTATTGG-----GGGGGTGGGGGAAA 1299  
QY 66 GlyPheGlyGlyLeuArgSerCysGlyLys----- 75  
Db 1298 GGTGGGGGTGATATTGTTAATGTGTGCTCCGCTTATGAGGGGGGTGGGGGGGG 1239  
QY 76 -----GlyTrpIleCys 79  
Db 1238 GTGGATATGAGATGGGGGGGTGACGTCCTCATATGACAGGGCGTGGATATTGTTG 1179  
QY 80 ProCys-----CysAlaGlyLysValGlyAlaHisArgAlaAsp 92  
Db 1178 CAGTGTGGCGAGATATGGGTATTATCGGGGGCGATTCCGA-----AGGTGAAG 1125  
QY 93 GluIleSerGlnValAlaHisGlnLeuGly-ThrGlySer-----ValAlaMetVal 110  
Db 1124 GAGCTCTTATGCTGTATATGCGGATATCGGCATATGAGGCGGTGATGATGAGCTCGT 1065  
QY 110 lThrMetThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTrpThrGlyLeu-- 129  
Db 1064 GACTATCGCT--AAACAT-----TTGTGGGTGGTGTGCTG 1032  
QY 130 -----SerAlaAlaLeuT-----Lys-AlaAlaThrAsnG 139  
Db 1031 CCGGACCGGGGAGTCACTGAGCGCTTATGAGGTTTGTGCGCGTGGGATATGCTGTG 972  
QY 139 lArgArGTTrpArgThrGluArgGluMetTyrGlyCysAspGlyTyrValArgAlaValG 159  
Db 971 GTAGCGGGCGCCCGGAGAGAGAGGAGTTAAAGTGAACGGG-----GCTTAT 921  
QY 159 lulleThrHisGlyLysAsnGlyTyrHisValHisValHisAlaLeuLeuMetPheSerG 179  
Db 920 GCTTACACTAGAGGCTTATAGAGGGGTGGCGAAGGGTGTGACGCTTTTGTGTTGTGG 861  
QY 179 lAspVal-----SerGluAsnIleLeuGluSerPheS 190  
Db 860 GCGGCGTGAAGATATGTCGCGGGTATAGAGAAATCCAGAAATTGGATGGCGCATATT 801  
QY 190 exAspAlaMetPheAspArgTyrThrSerLysLeuValSerLeuGlyPheAlaAlaProL 210  
Db 800 CGAACAGTAT-----CTCAGCCTTATATATGCG-----G 771  
QY 210 euArgAsnSerGlyGlyLeu-----AspValA 219  
Db 770 TAGCGGAATCAGCGCGCTGTAAATGATCATTAAGGTTGCCGANAAGTCCATGAGTGTG 711  
QY 219 rGlySerIleGlyGly-----GluAlaAspGlnValLeuAla 231  
Db 710 GCGTATAGAGTGTATTGTTGCCAGTATATGAGAGTGGAGAGGTATTAAGAGATGGGGA 651

```

QY      231 laTyrLeuThrLysIleAlaSerGlyValGlyMet----- 242
Db      650 GTATAGGTGCTGTTAGTAATCAAGGAGAGATTGGGTAGCGAGTACGTAAATGGG 591
QY      243 -----GluValGlySerGlyAspGlyLysSerGlyAspGlyAsnArgAlaProTyrG 261
Db      590 CCGAGGGGCCAGGGGTTTGGCGAGGGAGATAGGGGTATGTGGGTGGCAGCGCTCCCTTTT 531
QY      261 luIleAlaValAspAlaValAlaGlyLysProGlnAlaLeuGluLeuTyrArgGluPheG 281
Db      530 GGCATATGGGTGCGGAGCGCTTGGT-----GTTGAGTGGCTCAAGGATTTTG 466
QY      281 lupheGlySerMetGlyArgArgAlaIleAlaIleTyrSerArgGly-----LeuA 297
Db      485 GGGAGAGGTAGTGGTGGAGTGGCTGTATTGGGGCAGTAAGAGGGGATGATGTTGGCGC 426
QY      297 rg-AlaArgAlaGlyLeuGlyValAlaGluLeuThrAspAlaGlnIleValGluGlnGlu 316
Db      425 GTGGGCGAGGGGGGGGGGAGGGGCG-----GCGGGGGGTGGTGGGGCGAGGGT--- 381
QY      317 SerAlaProValMetValAlaIleIleProAlaArgSerTyrMetMetIleArgThrCys 336
Db      380 -----CCGGGGCGTGGGTGGTGGTCTTACTAGAGCTTTGT 348

RESULT 8
LOCUS   BI224374 1687 bp mRNA linear EST 11-JUL-2001
DEFINITION 602940916F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5104346 5',
VERSION BI224374
KEYWORDS BI224374.1 GI:14677818
SOURCE  Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1687)
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: egads-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLM at:
http://image.llnl.gov
Plate: L1AM11251 row: 0 column: 03
High quality sequence stop: 195.
Location/Qualifiers
FEATURES
Source
1..1687
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5104346"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_id="NIH_MGC_12"
/note="Organ: Cervix; Vector: PCMV-SHORT6; Site 1: NCI;
Site 2: Salt; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 286 a 842 c 254 g 305 t
ORIGIN

```

```

Query Match: 5.94% Indels: 124
DB: 12 Gaps: 18
US-10-007-527A-2 (1-379) x BI224374 (1-1687)

QY      61 GlyProLysGlySerGlyPheGlyLeuArgSerCysGlyLysGly-----Tyr 77
Db      1036 GGGAGGGTGGAGGTGGAGCGAGCGAGTGG-----GGCTGGCGAGAGCGCTGCAGGTGG 983
QY      78 IleCys-----ProCysValAlaGlyLys-----ValGlyAlaHisArg 90
Db      982 CTGTGATAGAGGGGCGAGTGTGTGGCGGAGCGCGCGCGCGAGTGGGTGGCCCGCG 923
QY      91 AlaAspGluIleSerGlnValAlaAlaHisGlnLeuGlyThrGlySerValAlaMetVal 110
Db      922 GCGCGAGAGGTG---AAGGAGGGGGGAGCGCGAGCGCGAAGGATAGAGAGCG----- 878
QY      111 ThrMetThrMetCargHisThrAlaGlyGlnArg---LeuHisAspLeuThrGlyLeu 129
Db      877 -----GCTGGAGCAGCGCTGAGAGAGAGAGTGTGGTGGGGGGCTG 839
QY      130 SerAlaAla-TryPheAlaAlaThrAsnGlyArgArgTyrArgThrGluArgGluMetC 149
Db      838 ACATGAGCACTGGCGGAGGGCGGAGAGCGGTGGCGCTGAGAGAGC----- 793
QY      149 rGlyCysAspGlyTyrValArgAlaValAlaGluIleThrHisGlyLysAsnGlyTyrHisVa 169
Db      792 -----AGGGGGTGGCGA-- 781
QY      169 HisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGluSer 189
Db      780 -----CATGCT----- 775
QY      189 SerAlaAlaMetPheAspArgTyrThrSerLysLeuValSerLeuGlyPheAlaAlaPr 209
Db      774 -----GGCAACGCCGCC 761
QY      209 OleuArgAsnSerGlyGly----- 215
Db      760 GCCCAGAGCAACCGCGGAGAGCGCGCGGGCGGAGAGAGAGAGAGAGAGTGGCG 701
QY      216 -----LeuAspValArgLysIleGlyGlyGlu-----AlaAspGlnValLeuAl 230
Db      700 GAGGCCAGTGTGGGACCGCGGCGCGGAGGCGGAGAGTGGGTGACAGTGGCGCG 641
QY      230 AlaTyrLeuThrLysIleAlaSerGlyValGlyMetGluValGlySer----- 246
Db      640 G-----GAGAGCGCGCGGGGTGTGACAGAGAGACCCGTGGAGAA 599
QY      247 ---GlyAspGlyLysSerGlyArgHisGlyAsnArgAlaProTyrGluIleAlaValAs 265
Db      598 GATGGGTGAGGCGGAATGAAATGACACAGCGAGCTGGCCGAGCGGAGGGTG 539
QY      265 AlaVal-----GlyGlyAspProGlnAlaLeuGluLeuTyrArgGluPhe 280
Db      538 GGTGACGAGAGGTGTGCGAGAGGGGCGACAGCGGGCGGAA-----TGGAGCGCGG 488
QY      280 eGluPheGlySerMetGlyArgArgAlaIleAlaIleTyr-----SerArgGlyLeuArgAl 298
Db      487 GCGCCGGGGGCGTGAAGGAGAGGGGGCGAGATGCTGTGACGACGAATGTGTGCCGC 428
QY      298 ArgAlaGlyLeuGlyValAlaGluLeuThrAsp-AlaGlnIleValGluGlnGluSerA 318
Db      427 AGCTAGGGGCGCATGGCGGAGATGCGCGAGGGCGCGAGACACGACAGGGCGAGAG 368
QY      318 laProValMetValAlaIleIle-----ProAlaArgSerTyrMetMetIleArgT 335
Db      367 CGCGCTGGGAGCGAGCGAGGCAAGGGCTGAGACTGCGAGGGCGCACTGTGTCCGGAG 308
QY      335 hrcysAlaProTyrValPheGlyGluIleLeuGlyLeuValGluAlaGlyAlaAlaThrTyr 354
Db      307 TGTGCAAGCGAGGGGTGCCAGCAGTGTCTGTGGTTCCATGGGGAATTAAGCTGTGG 249

```

Alignment Scores:

Pred. No.:	1.94	Length:	1687
Score:	119.00	Matches:	81
Percent Similarity:	33.72%	Conservative:	34
Best Local Similarity:	23.75%	Mismatches:	102

RESULT 9  
 LOCUS BQ714786/c 851 bp mRNA linear EST 16-UTL-2002  
 DEFINITION AGENCOURT 8482181 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6307476  
 5', mRNA sequence.  
 ACCESSION BQ714786  
 VERSION BQ714786.1 GI:21853685  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
 Tissue Procurement: Susan L. Sullivan, PhD.  
 cDNA Library Preparation: Resgen, Invitrogen Corp.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHM3725 row: a column: 13  
 High quality sequence stop: 666.  
 Location/Qualifiers  
 1..851  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6307476"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 129"  
 /note="Organ: olfactory epithelium; Vector:  
 pCMV-SPORT6.1.cdb; Site\_1: EcoRV, Site\_2: NotI; Cloned  
 unidirectionally. Primer: Oligo dt. Average insert size  
 2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this  
 is a NIH MGC Library."  
 BASE COUNT 164 a 289 c 234 g 164 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.964 Length: 851  
 Score: 118.00 Matches: 78  
 Percent Similarity: 32.43% Conservative: 30  
 Best Local Similarity: 23.42% Mismatches: 112  
 Query Match: 5.89% Indels: 113  
 DB: 13 Gaps: 15  
 US-10-007-527a-2 (1-379) x BQ714786 (1-851)  
 QY 68 GYGYLYEUAIGSERCYSGLYLYSGLYTRPIECYSPROCYSCYALAGLYLYVALGly 87  
 DB 845 GGGGGAAGCGGATACGACCGTGGGGATAC-----ACCGGTTGATTG 801  
 QY 88 ALAHISARGALASPGILLESERGINVALALAHISGLNLEU----- 102  
 DB 800 GCCCAATGGCG-----CTCTCAAAAGCGCTGGCCACATCTCTGACCCCTTGCT 747  
 QY 103 -----GLYThrGlySerValAlaMetValThrMetThrMetArgHisThr----- 117  
 DB 746 GACTTTGGCAAGGTCACAGTCTCAGAGCGATTGTGGAGACATGCTCAATGATGTCAT 687  
 QY 118 ALAGLYGlnArgLeu-----HisAspLeuTrpThrGlyLeuSerAlaAlaTrpLysAla 135  
 DB 666 GGTGGGAACATCTCACTAGACCTGATTACTTTGAGGTAGACAGGCGCTGGAGACC 627  
 QY 136 -----AlaThrAsnGlyArgArgTrpArgThrGluArgGluMet 148  
 DB 626 AGGTGAGCCAGAGCATGAGACGGCGGTGAGGAGAGGCTTCTGACACATGTTGCTC 567

QY 149 TyrGlyCysAspGlyTrpValArgAlaValGluIleThrHisGlyLysAsnGlyTrpHis 168  
 DB 566 TGTGGGACCTTGGAATC----- 543  
 QY 169 VALHISVALHISALALEUleuMetPheSerGlyAspValSerGluAsnIleLeuGluSer 188  
 DB 542 ACTAGGCTCCACAGATCTCTTGACCTCGGC----- 510  
 QY 189 PheSerAspAlaMetPheAspArgTrpThrSerIleValSerLeuGlyPheAlaAla 208  
 DB 509 -----TCTCAGGCGCCGTGCTGGGCTCCCAACT 480  
 QY 209 ProLeuArgAsnSerGlyLysAspValArgLysIleGlyGlyGluAlaAspGlnVal 228  
 DB 479 CCGTCCCGCTCAGCAGCACTGAGCTCAGA-----GACCGAGTC 438  
 QY 229 LeuAlaAlaTrpLeuThrIleValLeuSerGlyVal-----GlyMetGluValGlySerGly 247  
 DB 437 CTCTGAGGGGATCTTGCCCGGCTGTGACCAACATTAACAGTTGATCTTGTTCGCTT 378  
 QY 248 AspGlyLysSerGlyArgHisGlyAsnArgAlaProTrpGluIleAlaValAspAlaVal 267  
 DB 377 GACGGGGATGGCAACGCGGGAGCCAGGGTCAGCTGAGGGCTTAGGGTCAGC----- 324  
 QY 268 GlyGlyAspProGlnAlaLeuGluLeuTrpArgGluPheGluPheGlySerMetGlyArg 287  
 DB 323 -----TGG-----AGG 318  
 QY 288 ArgAlaIleAlaTrpSerArgGlyLeuArgAlaArgAlaGlyLeuGlyValGluLeuThr 307  
 DB 317 GCTGGGCTCAGCTGGGGAACGTGGCTCAGCAGAGAACATGCGGCGAGG----- 270  
 QY 308 AspAlaGlnIleValGluGlnGluGluSerAlaProValMetValAlaIleIleProAla 327  
 DB 269 -----GTCCATTCAGCGACGCGTAACCT-----CCGGCG 240  
 QY 328 ArgSerTrpMetMetIleArgThrCysAlaProIleValPheGlyGluIleGluGlyLeu 347  
 DB 239 CGC-----GAGCGGCGCGGGCTG 222  
 QY 348 ValGluAlaGlyAlaThrTrp-GluAsnLeuArgAspHisIleHisIleTrpArgLeuProAl 367  
 DB 221 GTCACTCCCGGCTTGGCTGGGGCGCGCGGCTTGGCGGCTGCTGCTCGGCTGAACGC 162  
 QY 367 AAlaAspValArgProProIleIleSerValArgLys 379  
 DB 161 AGGAGCCAGAGGCGCACCGGCTCGGGGTAGACAG 125  
 RESULT 10  
 LOCUS BZ554247/c 893 bp DNA linear GSS 17-DEC-2002  
 DEFINITION pasc1-60\_4590.y2 pasc1-60 Pseudomonas aeruginosa genomic clone  
 pasc1-60\_4590, genomic survey sequence.  
 ACCESSION BZ554247  
 VERSION BZ554247.1 GI:27161340  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 893)  
 Spencer,D.H., Raymond,C.K., Smith,B.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 Whole-Genome-Sequence Variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 J. Bacteriol. (2002) In press  
 CONTACT: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: [craymond@u.washington.edu](mailto:craymond@u.washington.edu)

FEATURES	Class: shotgun.	Location/Qualifiers
source	1.. 893	/organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="1-60" /db_xref="taxon:287" /clone="pacsi-60 4590" /clone_lib="pacsi-60" /note="Clinical isolate 1-60 whole genomic shotgun library."
BASE COUNT	151 a	332 c 270 g 138 t 2 others
ORIGIN		
Alignment Scores:		
Pred. No.:	1.03	Length: 893
Score:	118..00	Matches: 74
Percent Similarity:	34.07%	Conservative: 19
Best Local Similarity:	27.11%	Mismatches: 116
Query Match:	5.89%	Indels: 65
DB:	29	Gaps: 10
US-10-007-527A-2 (1-379) x BZ554247 (1-893)		
Qy	16	ProValleuValSerSerAspLysArgGlyLeuArgHisgluLeuArgProLysLeuGln 35
Db	730	CGAGTCTCTGTCCGGCGAACCCTTCGGGGCGTTCAATTCGGGAC-AGCCAGAAAGCCAG 672
Qy	36	GlnIleThrThrSer-----GluThrPheAsnAlaCysGlyArgProIle 50
Db	671	CAGCGCAGCCCGCAGATGCTCGGCAGTCTCGGCTTCACGCGCGTGAACGGCGCGCG 612
Qy	51	SerGlyValasnGlyValThrIleValasnGlyProLysGlySerGlyPheGlyGlyLeu 70
Db	611	GTTGGC-----GTAGNCGCCAGACAGCGCGCGCGCGCGCGCGAGCGGAGTC 570
Qy	71	ArgSerCysGlyLysGlyYTrpIleCysProCysCysAlaGlyLysValGlyAlaHisArg 90
Db	569	CAGNCT-----GCCGTGGCACTTGGAGCACCAAGTC 540
Qy	91	AlaAspGlnIleSerGlnValValAlaHisGlnLeuGlyYThrGlySerValAlaMetVal 110
Db	539	GTTCAAGGCGATCGCCAGGTCGTAGAGCATCCACCGGAGCAGGC--GTTGTAAGAATC 483
Qy	111	ThrMetThrMetArgHisThrAlaGlyGlnArgLeuHisAspLeuTrpThrGlyLeuSer 130
Db	482	GATCAGACCGCGCAGGTGCGGGCGCGCGCAACAGCA-CGTATTCGCGGAACAGTCCGCAT 424
Qy	131	AlaAlaIleTrpLysAlaAlaIleThrAsnGlyArgArgTrpArgThrGlnuArgLumeCtyrGly 150
Db	423	GGAAGGTGGCGCGCGCGCAGCGCAGGGCGCGCTCCGCATCCACCGCGCGATTCGCCCAAGG 364
Qy	151	CysAspGlyTyrValArgAla--ValGlnIleThrHisGlyLysAsnGlyYTrpHisVal 169
Db	363	CCGGGGCGCAGCAGGGCGCGCGCGCTGTTCCGGTATGTCGGCGGGGA---GGTTGGCAC-- 310
Qy	170	HisValHisAlaLeuIleuMetPheSerGlyAspValSerGluAsnIleLeuGluSerPhe 189
Db	309	-----CTTGTTCACGA 238
Qy	190	SerAspAlaMetPheAspArgTrpThrSerLysLeuValSerLeuGlyPheAla----- 207
Db	297	TCCAGGCGACAGCACCGGTGCTGGCGGCGTTCCAGAGATGCCCGCGGTTGCCGCTGAA 238
Qy	208	-----AlaProLeuArgAsnSerGlyGlyLeuAspValArgLysIleGly 222
Db	237	GGTGGCCGACAGATGGCCCATCTTCCTGGGAGTGTGGCGTTCCGCTGCCGCTCGAGC 178
Qy	223	GlyGlnAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaSerGlyValGlyMet 242
Db	177	GGCCGGCCAGCGCGGTTGCAGCAGCG-----CCGGCTTGCTT 139
Qy	243	GluValGlySerCysLysArgGlyLysSerGlyArgHisGlyAsnArgAlaProTrpGlnIle 266

Db	138	CGAGACGGCCACGGGCTCCGCGCTCCGGGGGCGCGACAGTACGAGCACC	-----	88
Qy	263	AlaValaspAlaValGlyGlyAspProGlnAlaLeuGln	275	
Db	87	-----GGGAGTCCACTAGTTCTTAGAG	67	
RESULT 11				
LOCUS	BQ936859/c			
DEFINITION	BQ936859	1306 bp	mRNA	linear
ACCESSION	AGENCOURT 10015321	NCI_CGAP Mam2	Mus musculus	CDNA clone
VERSION	IMAGE:6487576 5'			mRNA sequence.
KEYWORDS	BQ936859			
SOURCE	BQ936859.1	GI:22352242		
ORGANISM	EST.			
	Mus musculus			(house mouse)
REFERENCE	Mus musculus			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
JOURNAL	1 (bases 1 to 1306)			
COMMENT	NIH-MGC http://mgs.nci.nih.gov/.			
	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgapdb-remail.nih.gov			
	Tissue Procurement: Gilbert Smith, Ph.D.			
	CDNA Library Preparation: Life Technologies, Inc.			
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LNLN at:			
	http://image.lnl.gov			
	Plate: LRAM14032	row: m	column: 17	
FEATURES				
source	High quality sequence stop: 125.			
	Location/Qualifiers			
	1..1306			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/strain="FVB/N-3"			
	/db_xref="taxon:10090"			
	/clone IMAGE:6487576"			
	/tissue_type="tumor, biopsy sample"			
	/dev_stage="5 months"			
	/lab_host="DH10B"			
	/clone_id="NCI_CGAP Mam2"			
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;			
	Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.			
	Library constructed by Life Technologies. Investigator			
	providing samples: Gilbert Smith, NIH"			
BASE COUNT	216 a	571 c	327 g	190 t
ORIGIN				2 others
Alignment Scores:				
Pred. No.:	2.1			1306
Score:	117.00			Matches: 91
Percent Similarity:	32.38%			Conservative: 22
Best Local Similarity:	26.07%			Mismatches: 100
Query Match:	5.84%			Indels: 137
DB:	13			Gaps: 19
US-10-007-527A-2 (1-379) x BQ936859 (1-1306)				
Qy	45	AlaCysGlyValGProIleSerGlyValAlaGlnGlyValThrIleValAsnGlyPro	-----	62
Db	947	GCATGTGGGCGA-----GCGGGGGTCTCTGTGGCGCGGTACGTACGGGCGCGCGCA	894	
Qy	63	-----LysGlySerGlyPheGlyGly-----LeuArgSerCys	73	
Db	893	TGGGCGCGGAACGATGGAAATGGGTGGATACGGCGGGAACGAGAGATACACAAGCC	834	
Qy	74	GlyLysGlyTyrIleCysProCysCysAlaGlyLysValGlyAlaHisArgAlaSerGln	93	
Db	833	GGG-----GTGGGTAAACCGCGAGCGGAT---	810	

OY	94	lleserGlnValValAlaHisGlnLeuGlyThrGly-----	SerValAlaMetVal	110
Db	809	-----	..ATGGGCGCGAAGCGGTGGTAAAGTCGGCGCATGTGA	774
OY	111	ThrMetThrMetArgHisThrValaglyGlnArgLeuHisAspLeuThrThrGlyLeuSer		130
Db	773	-----	CGGGATACGGGCGGAGATGGCGCGCAT-----	ATCGGC 741
OY	131	AlaAlaTrpLysAlaAlaAlaThrAsnGlyArgArgTrpArgThrGluAlaGluMetTrpGly		150
Db	740	CGCGCGCGCTACCCGGAGATGTGTCCGGCGCGCGCTTGGAGATATGGGGTACTCCGATTAACGG		681
OY	151	CysAspGlyTrpValArgAlaValGlnLeuThrHisGlyLysAsnGlyTrpHisValHis		170
Db	680	-----	GGGNG-----	675
OY	171	ValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGlnSerPheSer		190
Db	674	-----	GGCGGGGTAAAG-----	663
OY	191	AspAlaMetPheAspArgTrpThrSerLysLeuValSerLeuGlyPheAlaAlaProLeu		210
Db	662	-----	CGTGTCCCGGTGCTGCTGCTGCGCGGGGGAGCGGGTTCGGCARG	618
OY	211	ArgAsnSerGlyGlyLeuAspValArgLysIleGlyGlyAlaAlaAspGlnValLeuAla		230
Db	617	GGACGCAATGGCGGCG-----	AGGCGGGGGGGGGGGGACTCAGCGCGCTCGTTGGG	567
OY	231	AlaTrpLeuThr-----	LysIleAlaSerGlyValGlyMetGluValGlySer---	246
Db	566	GCGCGCGCGAAGATGGGGGGCGCATTCGGGGGGCGGTGGATGGCGCGCGCGCTTCGGGC		507
OY	247	-----	GlyAspGlyLysSerGly-----	ArgHisGlyAsnArg 257
Db	506	ATCGGGAGGGGAACAGCGCGACCGCTCCGGGGGGGGGGCGGAGATCGCGCGGGCGCGGG		447
OY	258	AlaProTrpGlnIleAlaValAspAlaValGlyGlyAspProGlnAlaLeuGlnLeuTrp		277
Db	446	GCTCCC-----	GGGAACGGAAG-CCGACGCGGTGCAGACACCG	409
OY	278	ArgGluPheGluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGly-----		295
Db	408	CGTGGCGCTCGGTGTTCCGGGTCGGGCGCCCTGCAGACGCTGTGGTTGGAGAGAGCGCA		349
OY	296	---LeuArgAlaArgAlaGlyLeuGlyValaglyLeuThrAspAlaGlnIleValGluGln		314
Db	348	TGGCCTTAAGATCGCGCGCGCTTCGTCGACAGAGGCGCTCCGTGGCGGCTCGACGCGAG		289
OY	315	GluGlnSer-----	-----	AlaPro 319
Db	288	TAGAGGGGGAACACTACAGATGGGTAGTCTGGGACAGGGCTAGCGCCACCTTGGACCCC		229
OY	320	Val-MetVal-----	-----	AlaIleIleProAlaArgSe 329
Db	228	TGTGTTTGTGTCTCTGAGGCTGAGTGCAGATAAGTCACTACTACTCCAGGGAAACG		169
OY	329	rTrpMetMetIleArgThrCysAla 337		
Db	168	TTGGTGTAGAAATCCGAAGCTGTTC 144		

RESULT 12  
 BF783863/c  
 LOCUS BF783863 1615 bp mRNA linear EST 12-JAN-2001  
 DEFINITION 602109787P1 NCI CGAP K1d14 Mus musculus cDNA clone IMAGE:4328036  
 5', mRNA sequence.  
 ACCESSION BF783863  
 VERSION BF783863.1 GI:12088899  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	1 (bases 1 to 1615)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MSC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L14M8848 row: n column: 21  
 High quality sequence stop: 160.

FEATURES	Location/Qualifiers
source	1. .1615

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4238036"
/lab_host="DH10B (71 phage-resistant)"
/clone_1b="NCI CGAP Kid14"
/note="Organ: kidney; Vector: pCMV-SPORT6; site_1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."
BASE COUNT
ORIGIN
311 a      984 c      205 g      115 t

```

Alignment Scores:	
Pred. No.:	2.79
Score:	117.00
Percent Similarity:	25.36
Best Local Similarity:	22.97%
Query Match:	5.84%
DB:	10
Gaps:	14
Length:	1616
Matches:	79
Conservative:	22
Mismatches:	112
Indels:	132
Gaps:	14

US-10-007-527A-2 (1-379) X BF783863 (1-1615)

[illegible]

QY		158	VaIGluIlethrHrHSGLylyAsnGlyTThPHisValHSVAlHSAleuLeuMetPhe	177
Db		1003	GTTGGGTGTGGAGTGTAAGCGGTGTGTGTGGT-GTG-----	963
QY		178	SerGIysAspValSerGIuaSnIIeLenuIsSerPheSerAaspAlmetPheAArgTrp	197
Db		963	-----	963
QY		198	ThrSerIysLeuValSerLeuGlyPheAlaAlProLeuAArgAsnSerglyGlyLeuAsp	217
Db		962	-----TTGGTGGGGGTGTGGGGGGGGCGTCGGGTGTGGCGGAGAGGGTGGGGGTGT---	915
QY		218	ValArgylsIlleglyGlyGluAlaAspGlnValLeu-----AlaAlATyrLeuThr	234
Db		914	-----GGGGTGGGCCCGCCGCCTGGTGTGTGTGGTCGGCGTGTGGGGTGGGGGTG	864
QY		235	LysIleAlaSerGIysValGIyMetGCluValGIySer-----GlyAspGly	249
Db		863	CGGGTCCGGGGTGGCCCGGGGGGGAGAGGTCTCGAGCGGGGTGGGGGGCCTGGGGGGGGGG	804
QY		250	LysSerGIysArgHSglYsglyAsnaArgAlaProTrpGluIleAlaVal-AspaLaValGIyGI	269
Db		803	AGGTCGGGGCGCGGGGGGTTCGGGGCGCGGTG-----GTGCGGGGGCGTGGGGGGG	753
QY		269	YAspProGlnAlaLeuGluLeuTrp-----Ar	278
Db		752	CGGGAG-----GAGAGGTGTGGGGCGGGGTGTGGGGCGCGGTGAGAGGGGGCG	702
QY		278	gGlUpHeGluUphEGlyserMetGIyArgArgAlaIleAlatrpser-----	293
Db		701	TGAAGGGCGGGGGGGGTAACTGGGGGGCGGACCGGGGGGCGTGGGGCGCCGGGGCGGGGGA	642
QY		294	-----ArgGlyLeuArgAlaArgAl	300
Db		641	GGGGGGTGTGTGTGTGTGTGTCGTGCGTGTGCGCGGAGAGCGTGGCGGGGGTTTGAATGCGCGTGTG	582
QY		300	aGlyLeuGly	303
Db		581	CGGGCGTGGG	572

RESULT 13  
BUT05875/c

LOCUS DEFINITION BUT05875 UI-M-FR0-cai-a-04-0-U.r1 NIH\_BMAP\_FR0 Mus musculus cDNA clone IMAGE: 6412299 5', mRNA sequence.

ACCESSION BUT05875

VERSION BUT05875.1 GI:23635790

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE NIH-MGC http://mgc.mci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished

AUTHORS Contact: Robert Strausberg, Ph.D.  
Email: rsra@bbs-remail.nih.gov

JOURNAL Tissue Procurement: Dr. Jin Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LML at:  
<http://image.llnl.gov>

COMMENT This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES  
Seq primer: PYX-5.  
Location/Qualifiers  
1..689  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6"

DB	Accession	Sequence	Length	Matches	Conservative	Mismatches	Indels	Gaps
Db	336	CGCTTGAAGGGGATGGCCAAACGGGGGAGGCCAGGGTCAAGCTGAGGGCTAGCGTACG- 278	669	80	22	88	127	17
Qy	246	ecgIyAspGlyysSerGlyarGhIsGlyAsmArGlaJarProTppGluIleAlaValAspa 266	669	80	22	88	127	17
Db	396	GAGTCCTCTGAGGGGATCTTGCCCGGCTGTTCAGCAACAATAACAGTTGACTCTTTGTT 337	669	80	22	88	127	17
Qy	227	InValIleuAlaAlaTyrlleuThrLysIleAlaSerGlyVal--GlyMetGluValGlyS 246	669	80	22	88	127	17
Db	438	AACTTCGCGCCCGCTCAGGACGAGACTGGACTCAGA-----GAGC 397	669	80	22	88	127	17
Qy	207	IaAlaProLeuArGAsnSerGlyGlyLeuAspValArGylsIleGlyGlyAlaAspG 227	669	80	22	88	127	17
Db	463	-----TCTCAGGGCCCGTCCGCGCTCC 439	669	80	22	88	127	17
Qy	167	rPhIsvAlaHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuG 187	669	80	22	88	127	17
Db	501	GGGGTACTAGGCTCCACAGACATCTCTTGACCTCGGGC----- 464	669	80	22	88	127	17
Qy	187	IuSerPheSerAspAlaMetPheAspArgTTrpThrSerLysIleuValSerIleuGlyPheA 207	669	80	22	88	127	17
Db	463	-----TCTCAGGGCCCGTCCGCGCTCC 439	669	80	22	88	127	17
Qy	73	CysGly-----LysGlyTrpIleCysProCysCysAlaGlyLysValGlyAlaHisArg 90	669	80	22	88	127	17
Db	668	TGTGAGCATGCTCCAAATGG-----TGTCCATCTGC----- 636	669	80	22	88	127	17
Qy	91	AlaAspGluIleSerGlnValValAlaHisGlnIleuGlyThrGlySerValAlaMetVal 110	669	80	22	88	127	17
Db	635	-----GGAACTTCCTCACTTACGACTGAT 612	669	80	22	88	127	17
Qy	111	ThrMetThrMetArGhIsThrAlaGlyIlnArGleuHisAspLeuTrpThrGlyLeuSer 130	669	80	22	88	127	17
Db	611	ACTTGTAGCGTA-----GGACAAAGCGCTGNGAGACCA-----GGCTTAGGCC 571	669	80	22	88	127	17
Qy	131	AlaAlaTrpLysAlaAlaThrAsnGlyArGArGTrp-ArgThrGln-----ArgGluMet 148	669	80	22	88	127	17
Db	570	AGAGGATGGAG-----CGCTGTAGCGCAGAGACTTCTTGACACAT 529	669	80	22	88	127	17
Qy	148	CTyrgIly-CysAspGlyTyTrValArGAlaValGluIleThrHisGlyLysAsnGly---T 167	669	80	22	88	127	17
Db	528	GSTTCTCTCT-----GGGCACTTCGATCT 502	669	80	22	88	127	17
Qy	167	rPhIsvAlaHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuG 187	669	80	22	88	127	17
Db	501	GGGGTACTAGGCTCCACAGACATCTCTTGACCTCGGGC----- 464	669	80	22	88	127	17
Qy	187	IuSerPheSerAspAlaMetPheAspArgTTrpThrSerLysIleuValSerIleuGlyPheA 207	669	80	22	88	127	17
Db	463	-----TCTCAGGGCCCGTCCGCGCTCC 439	669	80	22	88	127	17
Qy	207	IaAlaProLeuArGAsnSerGlyGlyLeuAspValArGylsIleGlyGlyAlaAspG 227	669	80	22	88	127	17
Db	438	AACTTCGCGCCCGCTCAGGACGAGACTGGACTCAGA-----GAGC 397	669	80	22	88	127	17
Qy	227	InValIleuAlaAlaTyrlleuThrLysIleAlaSerGlyVal--GlyMetGluValGlyS 246	669	80	22	88	127	17
Db	396	GAGTCCTCTGAGGGGATCTTGCCCGGCTGTTCAGCAACAATAACAGTTGACTCTTTGTT 337	669	80	22	88	127	17
Qy	246	ecgIyAspGlyysSerGlyarGhIsGlyAsmArGlaJarProTppGluIleAlaValAspa 266	669	80	22	88	127	17
Db	336	CGCTTGAAGGGGATGGCCAAACGGGGGAGGCCAGGGTCAAGCTGAGGGCTAGCGTACG- 278	669	80	22	88	127	17

QY 266 laValGlyAspProGlnAlaLeuGluLeuTrpArgGluPheGluPheGlySerMetG 286  
 Db 277 -----TGG----- 275  
 QY 286 lYArgAlaAlaIleAlaTrpSerArgGlyLeuArgAlaArgAlaGlyAlaGlu 306  
 Db 274 --AGGGCTGGCTCAGCTGGGGAACCTGGCTCAGGAGAACTGGGGCAGGA----- 224  
 QY 306 eutHrAspAlaGlnIleValGluGlnGluSerAlaProValMetValAlaIleLeu 326  
 Db 223 -----GTCCAAATCCAGCAGCCGTAACT-----C 199  
 QY 326 roAlaArgSerTrpMetMetIleArgThrCysAlaProTyrValPheGlyGluIleLeu 346  
 Db 198 CGGCGGCGC-----GAGCGGCGCG 181  
 QY 346 lYLeuValGluAlaGlyAlaThrTrp-GluAsnLeuArgAspHisLeuHisTyrArgLeu 365  
 Db 180 GGGTGGTCACTCCCGGGTTGGGCTGGGGGCGCGGCGCTTGGGCTGCTCGCGGGT 121  
 QY 366 ProAlaAlaAspValArgProProIleIleSerValArgLys 379  
 Db 120 AACGACAGAGCCACAGGCGCACGAGCTCGGGGTGAGACAG 79  
 RESULT 14  
 B1411296/c 867 bp mRNA linear EST 14-AUG-2001  
 LOCUS 602964678F1 NCI\_CGAP\_Lu33 Mus musculus cDNA clone IMAGE:5120034 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1411296  
 VERSION B1411296.1 GI:15172219  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 867)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LIML at:  
 http://image.jiml.gov  
 Plate: LHAM11292 row: 1 column: 19  
 High quality sequence start: 32  
 High quality sequence stop: 861.  
 Location/Qualifiers  
 1..867  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="CZECH 11"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5120034"  
 /tissue\_type="pooled lung tumors"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NCI CGAP Lu33"  
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site 1: NotI; Site 2: EcoRI; 1st  
 strand cDNA was prepared from mRNA obtained from pooled  
 lung tumors with a Not I - oligo(dT) primer [5'.  
 TGTACCAATCTGAAGTGGAGCGCGCTGTGTTTCTTTTCTTTT 3'].  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not  
 I and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was

BASE COUNT 171 a 292 c 234 g 170 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.51 Length: 867  
 Score: 116.00 Matches: 83  
 Percent Similarity: 32.31% Conservative: 22  
 Best Local Similarity: 25.54% Mismatches: 89  
 Query Match: 5.79% Indels: 132  
 Db: 12 Gaps: 18  
 US-10-007-527a-2 (1-379) x B1411296 (1-867)  
 QY 69 GlyLeuArgSer-----CysGly-----LysGlyTrpIleCysProCys 81  
 Db 710 GGCAGAGGTCACAGCTTCAGACGATGTGGAGCATGCTCCAAATAG--TGTCCATGG 654  
 QY 82 CysAlaGlyLysValGlyAlaHisArgAlaAspGluIleSerGlnValAlaHisGln 101  
 Db 653 TGC----- 651  
 QY 102 LeuGlyThrGlySerValAlaMetValThrMetThrMetArgHisThrAlaGlyGlnArg 121  
 Db 650 ---GGACATCTCACTACACCTGATCTGTAAGGTA-----GGACAAAGC 606  
 QY 122 LeuHisAspLeuTrpThrGlyLeuSerAlaAlaTrpLysAlaAlaThrAsnGlyArg 141  
 Db 605 CTGGAGACCA-----GGCTGAGCCAGAGCATGTAGAG-----CGC 571  
 QY 142 Trp-ArgThrGlu-----ArgGluMetCysGlyTyrValArgAlaValG 159  
 Db 570 TGGTAGGACGAGAGCTTCTTCAACATGTTGCTCTG----- 533  
 QY 159 IuIleThrIleGlyLysAsnGly---TrpHisValHisValHisAlaLeuMetPheS 178  
 Db 532 -----GGGACCTTGGCATTTGGGATGAGTACAGCTCCACAGCATTTCTTGGACT 484  
 QY 178 eArgLysAspValSerGluAsnIleLeuGluSerPheSerAspAlaMetPheAspArgTrp 198  
 Db 483 CGGGC----- 479  
 QY 198 HisSerLysLeuValSerLeuGlyPheAlaAlaProLeuArgAsnSerGlyLysLeuAsp 218  
 Db 478 --TTCAGAGCCCGTGGCTGCTCCCACTTCCGGCCCGCTCAGCAGCAGCATGAGCA 421  
 QY 218 AlaArgLysIleGlyGlyAlaAspGlnValLeuAlaAlaTyrLeuThrLysIleAlaS 238  
 Db 420 TCGAG-----GACGAGTCTCTGAGGGAGATCTGCCCGGCTTTGA 379  
 QY 238 eArgLysVal---GlyMetGluValGlySerGlyAspGlyLysSerGlyArgHisGlyAsn 257  
 Db 378 CCAACATAAACAGTTGTACTCTTGTGCTTGACGAGGGAGATCGGCAACGGGGAGCCA 319  
 QY 257 ArgAlaProTrpGluIleAlaValAlaAspAlaValGlyLysProGlnAlaLeuGluLeu 277  
 Db 318 GGGTCAGCTGAGAGGCTAGGTCAGC-----T 292  
 QY 277 rPArgGluPheGluPheGlySerMetGlyArgArgAlaIleAlaTrpSerArgGlyLeu 297  
 Db 291 GG-----AGGGCTGGCTCAGCTGGGAACTGGAGCTCA 259  
 QY 297 rGAlaArgAlaGlyLeuGlyAlaGluLeuThrAspAlaGlnIleValGluGlnGlu 317  
 Db 258 GGCAGAGACTGGGGCAGGA-----GTCCAAATCCAGGACAGC 223  
 QY 317 eAlaProValMetValAlaIleIleProAlaArgSerTrpMetMetIleArgThrCys 337  
 Db 222 CGTAACCT-----CCGGCGCGC----- 206  
 QY 337 lAProTyrValPheGlyGluIleLeuGlyLeuValGluAlaGlyAlaThrTrp-GluAsn 356  
 Db 205 -----GAGCGGCGCGGCTGCTACCTCCGGGTTCCGCTGGGGCCGG 163



QY 357 LeuArgAspHisIleuHisIstYrArGleuProAlaAlaAspValArgProIleIleSer 376  
Db 162 GGCGCTTGCGGCTGCTGCTGCGGTGAACGACGAGCCACGAGCGGCTCGG 103  
QY 377 ValArgLys 379  
Db 102 GTGAGACAG 94

RESULT 15  
BG684181/c  
LOCUS 602635811F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4763523 5',  
DEFINITION mRNA sequence.  
ACCESSION BG684181  
VERSION BG684181.1 GI:13915578  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LIML at:  
http://image.llnl.gov  
Plate: LCM1619 row: f column: 04  
High quality sequence start: 51  
High quality sequence stop: 147.  
Location/Qualifiers  
1. 1199  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4763523"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site: 1: XhoI;  
Site: 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGACAGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

BASE COUNT 263 a 502 c 300 g 134 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2.32 Length: 1199  
Score: 116.00 Matches: 75  
Percent Similarity: 31.53% Conservative: 24  
Best Local Similarity: 23.89% Mismatches: 111  
Query Match: 5.79% Indels: 104  
DB: 10 Gaps: 17

US-10-007-527a-2 (1-379) x BG684181 (1-1199)

QY 47 GLAArgProIleSerGlyValaAsnGlyValThrIleValaAsnGlyProLys---GlySer 65  
Db 1074 GGCGCGCGCGCGGCTGCGGTGAACGACGAGCCACGAGCGGCTCGG 1015  
QY 66 GlyPheGlyLys-----LeuArgSerCysGlyLysGlyTyrIleCys 79

Db 1014 GGCTGCGGCGGCTGCGGTGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCT 955  
QY 80 ProCysCysAlaGlyLysValGlyValaHisArgAlaAspGluIleSerGluVala 99  
Db 954 TTGCACTGTTGCGGCGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 895  
QY 100 -----HisGluLeuGlyThrGlySerVal 107  
Db 894 GCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 838  
QY 108 AlaMetValThrMetThrMetArgHisThrAlaGlyGluArgLeuHisAspLeuTyrThr 127  
Db 837 GCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 814  
QY 128 GlyLeuSerAlaAlaTyrLysAlaAlaThrAsnGlyArgArgTyrArgThrGluArgGlu 147  
Db 813 GGTGCTCTCGACGTGGGCGCATGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 760  
QY 148 MetTyrGlyCysAspGlyTyrValArgAlaValGluIleThrHisGlyLysAsnGlyTyr 167  
Db 759 TTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 718  
QY 168 HisValHisValHisAlaLeuLeuMetPheSerGlyAspValSerGluAsnIleLeuGlu 187  
Db 717 -----CATGCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 685  
QY 188 SerPheSerAspAlaMetPheAspArgTyrThrSerGlyLysValSerLeuGlyPheAla 207  
Db 684 GGCACCTGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 625  
QY 208 AlaProLeuArgAsnSerGlyLysLeuAspValArgLysIleGlyGluAlaAspGlu 227  
Db 624 TGTCCG----- 619  
QY 228 ValLeuAlaAlaTyrLeuThrTyrIleAlaSerGlyValGlyMetGluValGlySerGly 247  
Db 618 -----GCTGGCATGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 595  
QY 248 -----AspGlyLysSerGlyValArgHisGlyAsn-----ArgAlaProTyr 260  
Db 594 CTTGTGTTATGTGTGGGAGGAGGTTGCAATGCACTGACCGCATGATGCGGCGG 535  
QY 261 GluIleAlaValaAspAlaValaGlyLysAspProGluAlaLeuGluLeuTyrArgGluPhe 280  
Db 534 TGTCTTCAAGAGAGAGCGGCTGACAGGTGACCACTGACGCTA----- 493  
QY 281 GluPheGlySerMetGlyArg-----ArgAlaIleAlaTyr-----SerArgGlyLeu 296  
Db 492 -----GCCGATGACAGAGCGGTGCTGCTGCGGTGTGTATGGCACCGGTGCGGCGGCTC 439  
QY 297 ArgAlaArgAlaGlyLeuGlyValaGluLeuThrAspAlaGluIleValaGluGluGluGlu 316  
Db 438 GAGGCGCGCGGTTTGCTCGGAGTACG-----GCCAGCTTGATTTGATGACAGAG 391  
QY 317 SerAlaPro-----ValMetValaAlaIleIleProAlaArg 328  
Db 390 TCCGCCAAGAGACGTTGCTTACGTGATGAGTGTCTTCACAGCGC 349

Search completed: August 22, 2003, 12:09:11  
Job time : 1932 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 21:40:18 ; Search time 60 Seconds  
(without alignments)  
1002.623 Million cell updates/sec

Title: US-10-007-527a-2

Sequence: 1 MTSVSAEHLGSKDRPPVTVS.....HHHRLPAADVPPITTSVRK 379

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_19Jun03:\*
- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
  - 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
  - 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
  - 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
  - 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
  - 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
  - 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
  - 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
  - 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
  - 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
  - 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
  - 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
  - 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
  - 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
  - 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
  - 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
  - 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
  - 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
  - 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
  - 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
  - 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
  - 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
  - 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*
  - 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2005	100.0	379	23	ABB84278 Rhodococcus AN12 r
2	609.5	30.4	459	23	ABB84280 A. pyogenes pAPI r
3	399.5	19.9	494	19	AAW73071 B. breve essential
4	326	16.3	427	21	AAW97536 B. lactofermentum
5	326	16.3	427	23	ABB06341 Brevibacterium lac
6	326	16.3	427	23	AAW13530 S. lactofermentum
7	322	16.1	451	23	ABB84283 S. nigificans PSN
8	318	15.9	427	21	AAW97537 B. lactofermentum
9	318	15.9	427	23	ABB06342 Brevibacterium lac

10	318	15.9	427	23	AAW13531
11	316	15.8	456	23	ABB84281
12	282	14.1	466	23	AAW11039
13	266.5	13.3	528	23	ABB84282
14	134	6.7	314	14	AAW93143
15	103	5.1	647	24	AAW56572
16	100.5	5.0	577	17	AAW06554
17	100	5.0	1568	22	ABG03109
18	97	4.8	3798	21	AAW58582
19	95.5	4.8	713	21	AAW58577
20	94.5	4.7	431	23	ABP65883
21	94.5	4.7	1721	19	AAW52847
22	94	4.7	386	22	AAE30185
23	93.5	4.7	356	22	ABE10262
24	93.5	4.7	356	23	ABP66849
25	93.5	4.7	1049	24	ABR41258
26	92.5	4.6	579	20	AAW35500
27	92	4.6	740	22	ABG24479
28	91.5	4.6	444	23	ABP66024
29	91.5	4.6	527	22	AAW59539
30	91.5	4.6	622	22	AAW71880
31	91	4.5	956	22	AAW36332
32	90.5	4.5	415	22	AAW41823
33	90.5	4.5	729	22	ABW59500
34	90	4.5	668	22	AAW63901
35	89.5	4.5	361	21	AAW40780
36	89.5	4.5	369	21	AAW40779
37	89.5	4.5	375	21	AAW40778
38	89.5	4.5	509	22	AAW67532
39	89.5	4.5	585	22	AAW92278
40	89	4.4	350	22	AAW61038
41	89	4.4	898	23	AAW09878
42	89	4.4	3562	22	AAW82213
43	89	4.4	3739	21	AAW18638
44	89	4.4	3739	21	AAW77193
45	89	4.4	3739	21	AAW77201

ALIGNMENTS

RESULT 1  
ABB84278 standard; Protein; 379 AA.  
XX ABB84278;  
AC ABB84278;  
XX 13-JAN-2003 (first entry)  
DT  
DE Rhodococcus AN12 replication protein Rep.  
XX  
XX  
XX Plasmid stability protein; replication protein; ethylene forming enzyme;  
XX carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
XX polyhydroxyalkanoic acid synthase; FFA synthase; nitrile hydratase;  
XX alcohol dehydrogenase; terpene synthase; cholesterol oxidase;  
XX shuttle vector; Rep protein.  
XX  
XX Rhodococcus erythropolis.  
XX  
XX W0200255709-A2.  
XX 18-JUN-2002;  
XX  
XX 12-DEC-2001; 2001WO-US47868.  
XX  
XX 12-DEC-2000; 2000US-254868P.  
XX  
XX (DUPO ) DU POINT DE MEMOURS & CO E I.  
XX  
XX Bramucci MG, Cheng Q, Kozlitchka KN, Tomb J;  
XX WPI, 2002-557827/59.  
XX  
XX N-PSDB; ABQ76122.

B. lactofermentum  
S. lividans p101  
Replication protei  
S. phaeochromoge  
Sequence encoded b  
Lung cancer associ  
R. capsulatus acet  
Novel human diago  
Sorangium cellulos  
Bifidobacterium lo  
A. mediterranei ri  
Human PHD1 protein  
Human CDNA SEQ ID  
Human polyprotein  
Human DITP intrac  
Chlamydia pneumonia  
Novel human diagno  
Bifidobacterium lo  
Propionibacterium  
Human PEPCK-cytoso  
Pseudomonas aerugi  
Propionibacterium  
Drosophila melanog  
Propionibacterium  
Zea mays protein f  
Zea mays protein f  
Propionibacterium  
C glutamicum prote  
Propionibacterium  
Novel human secret  
Polyketide synthas  
Amino acid sequenc  
S. venezuelae pik

XX New nucleic acid molecule encoding replication protein/plasmid  
PT stability protein, useful in cloning and expression vectors,  
PT particularly shuttle vectors for expression of heterologous genes in  
PT Rhodococcus species -  
XX  
XX Claim 4; Page 64-65; 96pp; English.  
XX  
XX This invention describes a novel nucleic acid encoding a replication  
CC protein or a plasmid stability protein. The product of the invention is  
CC useful for expression of nucleic acid such as genes encoding enzymes  
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
CC actinomycetales bacteria. The replication protein or plasmid stability  
CC protein are useful in cloning and expression vectors and particularly in  
CC shuttle vectors for the expression of homologous and heterologous genes  
CC in Rhodococcus sp. This sequence represents the Rhodococcus AN12 Rep  
CC protein described in the disclosure of the invention.  
XX  
XX  
SQ Sequence 379 AA;  
Query Match 100.0%; Score 2005; DB 23; Length 379;  
Best Local Similarity 100.0%; Pred. No. 1.5e-190;  
Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTSVSAEHLGSKRPPVLYSSDRGRHRELRLQOITTSFTNACGRPISGVNGVTVN 60  
DB 1 MTSVSAEHLGSKRPPVLYSSDRGRHRELRLQOITTSFTNACGRPISGVNGVTVN 60  
QY 61 GPKSGFGGLRSCGKMWICPCAGKVGARADEISQVVAHQLTGGSVAWVTMTMTAQQ 120  
DB 61 GPKSGFGGLRSCGKMWICPCAGKVGARADEISQVVAHQLTGGSVAWVTMTMTAQQ 120  
QY 121 RLHDLMTGLSAAMKAATNGRRRTERTERTYGGDYVRAVEITHGKNGHVTALLMPSGD 180  
DB 121 RLHDLMTGLSAAMKAATNGRRRTERTERTYGGDYVRAVEITHGKNGHVTALLMPSGD 180  
QY 121 RLHDLMTGLSAAMKAATNGRRRTERTERTYGGDYVRAVEITHGKNGHVTALLMPSGD 180  
DB 121 RLHDLMTGLSAAMKAATNGRRRTERTERTYGGDYVRAVEITHGKNGHVTALLMPSGD 180  
QY 181 VSNITLESFSDAMPDRMTSLVSLGFAAPLRNSGGIDVRKIGEDAOVLAAYLTAKASGV 240  
DB 181 VSNITLESFSDAMPDRMTSLVSLGFAAPLRNSGGIDVRKIGEDAOVLAAYLTAKASGV 240  
QY 241 GMEVSGDGKSGRRNAPWEIIVDAVAGDPQALREPERFSGMRRAIAMSRLRARA 300  
DB 241 GMEVSGDGKSGRRNAPWEIIVDAVAGDPQALREPERFSGMRRAIAMSRLRARA 300  
QY 301 GLGAEILTDAQIVEQESAPVWVAIIPARSMWMTIRTCAPYFGEILGLVEAGATWENLRDH 360  
DB 301 GLGAEILTDAQIVEQESAPVWVAIIPARSMWMTIRTCAPYFGEILGLVEAGATWENLRDH 360  
QY 361 LHYRLPADVRRPPIISVRK 379  
DB 361 LHYRLPADVRRPPIISVRK 379  
RESULT 2  
ABB84280  
ID ABB84280 standard; Protein; 459 AA.  
XX  
XX ABB84280;  
XX  
XX 13-JAN-2003 (first entry)  
XX  
DE A. pyogenes pAP1 replication protein.  
XX  
XX Plasmid stability protein; replication protein; ethylene forming enzyme;  
XX carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;  
XX polyhydroxyalkanoic acid synthase; PHA synthase; nitrile hydratase;  
XX alcohol dehydrogenase; terpene synthase; cholesterol oxidase;  
XX shuttle vector.  
XX  
OS Arcanobacterium pyogenes.

XX  
PN WO200255709-A2.  
XX  
XX 18-JUL-2002.  
XX  
XX 12-DEC-2001; 2001WO-US47868.  
XX  
XX 12-DEC-2000; 2000US-254868P.  
XX  
XX (DUPO ) DU PONT DE NEMOURS & CO E. I.  
XX  
XX Brannucci MG, Cheng Q, Kostichka KN, Tomb J;  
XX WPI; 2002-557827/59.  
XX  
XX  
XX New nucleic acid molecule encoding replication protein/plasmid  
PT stability protein, useful in cloning and expression vectors,  
PT particularly shuttle vectors for expression of heterologous genes in  
PT Rhodococcus species -  
XX  
XX  
XX Example 5; Fig 4A; 96pp; English.

CC This invention describes a novel nucleic acid encoding a replication  
CC protein or a plasmid stability protein. The product of the invention is  
CC useful for expression of nucleic acid such as genes encoding enzymes  
CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic  
CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile  
CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol  
CC dehydrogenase, terpene synthases, and cholesterol oxidase in an  
CC actinomycetales bacteria. The replication protein or plasmid stability  
CC protein are useful in cloning and expression vectors and particularly in  
CC shuttle vectors for the expression of homologous and heterologous genes  
CC in Rhodococcus sp. This sequence represents A. pyogenes pAP1 replication  
CC protein described in the disclosure of the invention.

SQ Sequence 459 AA;  
Query Match 30.4%; Score 609.5; DB 23; Length 459;  
Best Local Similarity 36.3%; Pred. No. 1.1e-51;  
Matches 148; Conservative 62; Mismatches 133; Indels 65; Gaps 15;  
QY 5 SAHELSGKRRPPVLYSSD-----KRGIRHRELRLQOITTSFTNACGR-PISGVNGV 56  
DB 67 SATHPPLGNTVLTFFPVSNESKRTAKSRRSRYELRDIAISTESVRRKGRVVAELVSL 126  
QY 57 TIYNGPKSGFGGLRSCGKMWICPCAGKVGARADEISQVVAHQLTGGSVAWVTMTMTA 115  
DB 127 RAKSDGKAGYGGELHTCGSVMACPVCSAKIARRKTDLOQVVDHAYKGNVTSMILITQR 186  
QY 116 HTAGORLHDLMTGLSAAMKAATNGRRRTERTERTYGGDYVRAVEITHGKNGHVTALL 175  
DB 187 HHKGGGLKHLMDALSTAMNRVTSGRRWIEFKQFGLVGVRAVEITHGKNGHVTALL 246  
QY 176 MBSGDVSENILESF-----SPAMDRMTSKVSLGFAAPLRNSGG 215  
DB 247 ISEKD---PLTSTFTVYQRRQGRRLPYPEIYNSDFIARWAGLAKGVDF-LDSGG 302  
QY 216 L-----DYRKIGEDAOVLAAYLTKI-----ASGVGMEVSGDGKSGRRNAPWEIIVDA 266  
DB 303 LDWTVAKDARAION-----YVSKMGSTDAISSEVLTGSKKARANKRTPFQILADI 354  
QY 267 VG-GDPQALREPERFSGMRRAIAMSRLRARAIGAEILTDAQIVEQESAPVWVAI 335  
DB 355 LSLGVDVDDIKLWKEYEKASFGRRALTWSKLRDMANLGVQSDDEIA-SPEIDDEAIALF 413  
QY 326 PARSMWMTIRTCAPYFGE--EILGLVAGA-----TWNLRDHLHYLTP 366  
DB 414 THDAMPQVRR-----FGAEELLDVTSGGRAAAYRWLDFRE-IDWSLP 455  
RESULT 3  
AAW73071  
ID AAW73071 standard; Protein; 494 AA.

```
XX AC AAW73071;
XX DT 05-JAN-1999 (first entry)
XX DE B. breve essential region gene protein.
XX KM Essential region gene; shuttle vector.
XX OS Bifidobacterium breve.
XX PN JP10262670-A.
XX PD 06-OCT-1998.
XX PF 27-MAR-1997; 97JP-0091387.
XX PR 27-MAR-1997; 97JP-0091387.
XX PA (HONS ) YAKULT HONSHA KK.
XX DR WPI; 1998-587288/50.
XX N-PSDB; AAV58945.
XX PT Shuttle vector for a Bifidobacterium species - contains specific
XX elements from existing plasmids e.g. pNBI of ATCC 15698
XX PS Claim 5; Page 14-15; 17pp; Japanese.
XX CC This sequence is encoded by the essential region gene from plasmid pNBI
XX used in the vector of the invention. The vector is a shuttle vector for a
XX Bifidobacterium, and contains: (a) a replicated essential region
XX originated from a plasmid pNBI of Bifidobacterium breve ATCC 15698;
XX (b) a replicated essential region originated from a plasmid of E. coli;
XX and (c) an antibiotic-resistant gene originated from a plasmid of E. coli
XX and an antibiotic-resistant gene functioning by a Bifidobacterium. The
XX shuttle vector can be used to transform various Bifidobacteria species.
XX SQ Sequence 494 AA;

Query Match 19.9%; Score 399.5; DB 19; Length 494;
Best Local Similarity 27.2%; Pred. No. 9,2e-31;
Matches 116; Conservative 64; Mismatches 169; Indels 77; Gaps 15;

QY 4 VSAEHLGSG-KDRP-----PVLVSSDKR-----GIRHELRLPKLQOITTSFPN----- 44
DB 38 VAAEAAASGARGSPPEKTKMTPTSLSRDLRLAYGRRAESKILVRHAGETTLGFEPIK 97
QY 45 -----ACGRPISGVNGVTIVNGPKSGFGGLRSCGKMTCPCCAGKVGARADEISQVVA 99
DB 98 LPRCARCGQPVDP--TGVGVMTNGEKARFTGTMLCGSIWACPTCSAIIIRERAHEVALAIG 155
QY 100 HQL-----GTGSVAWMTMTMRTAQRHLDMTGLSA 131
DB 156 NHAETLRRAAADQMAHEGQRLPELWVSDSFGVITFQTLRLDRTPMLMTLDALIK 215
QY 132 AWAATNGRWRTEREMWCGDGYRAVEITTHGKNGHVAHALLMPSGDVSENILIESFSD 191
DB 216 GHTKAINSGSPWQRAEERKIRGFVRAIETTYGVNGHPIIHVMPLDGLDLDGQGBAMQ 275
QY 192 AMFDEWTS--KLVSIGF-----AABLRSNGLDYR-KIGAGADQVLAAYLTIKIASG 239
DB 276 WLDLRWKTWKAAYAKKKKQDGNPVYVAP-NDEHGIDLOFKSGKAGTAAYITIKIQD 334
QY 240 VG-----MEVSGDDGSGHGRNARAPMEIANDAVG--DPQALMLREFPFGSGRRALAY 292
DB 335 KGGVTLAQBIAIAGDIKNGRMGVSVPQLDSDGCLGSDPQRBDLMEVWQATLRRCITW 394
QY 293 SRGLARAGLGAELTDAQIVQESAPVVA--IIPARSMMTIRTCAPYFGEIIGVEXG 351
DB 395 SRGLKEDMEV--ELEDEBELAKADELPGLVGVYVVRNVYKDIRKSAPELADALDAER- 452
QY 352 ATWENTL 357
```

```
DB 453 EDWQEV 458

RESULT 4
ID AAY97536 standard; Protein; 427 AA.
XX AAY97536;
XX AC AAY97536;
XX DT 15-JAN-2001 (first entry)
XX DE B. lactofermentum p48K protein sequence.
XX KM Temperature sensitive plasmid; TSCRCR; protein production;
XX temperature sensitive replication control region; p48K.
XX OS Brevibacterium lactofermentum.
XX PN EP1038966-A1.
XX PD 27-SEP-2000.
XX PF 16-MAR-2000; 2000EP-0105326.
XX PR 16-MAR-1999; 99JP-0069896.
XX PA (AJIN ) AJINOMOTO CO INC.
XX PI Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatsu T;
XX DR WPI; 2000-573832/54.
XX N-PSDB; AAA90934.
XX PT Plasmids containing a temperature sensitive replication control regions
XX useful for breeding microorganisms for the production of amino acids by
XX fermentation -
XX PS Claim 8; Page 14-17; 29pp; English.
XX CC This sequence is the Brevibacterium lactofermentum p48K protein.
XX The invention relates to a plasmid containing a temperature sensitive
XX replication control region (TSCR) and a marker gene (Mg). The TSCR is
XX derived from plasmid pM4330 harboured by Brevibacterium lactofermentum
XX (ATCC 13869) and allows the plasmid to replicate autonomously at a low
XX temperature but does not allow the plasmid to replicate autonomously at
XX an elevated temperature in coryneform bacteria within a temperature
XX range in which the bacteria can grow. The plasmid can be used for
XX modifying a chromosomal gene in a coryneform bacterium, which may be used
XX for the production of useful substances, such as amino acids, by
XX fermentation, to change their genetic traits. Therefore, the plasmid can
XX be used for breeding microorganisms for the production of amino acids by
XX fermentation. The plasmid comprises a TSCR that allows the plasmid to
XX replicate autonomously at an elevated temperature in coryneform
XX bacteria within a temperature range in which the bacteria can grow.
XX SQ Sequence 427 AA;

Query Match 16.3%; Score 326; DB 21; Length 427;
Best Local Similarity 29.1%; Pred. No. 1.5e-23;
Matches 98; Conservative 58; Mismatches 131; Indels 50; Gaps 15;

QY 34 LQOITTSFPNACGR-----PIGVNGVTIVNGPKSGFGGLRSCGKMTCPCCAGV 86
DB 27 MKITNSKALAGCHRRNRDEAVAYWS-----KG--ASQFGLNHSRMSPLAELEV 79
QY 87 GAHRADEISQVVAHQIGTSVAVTM-TMRTAQRHLDMTGLSAWKAATNGRWRTE 145
DB 80 MEERRIELAITKHLAAGALMVFVGRHNRBSFQVEXGITAAYASVWTKSQWKE 139
QY 146 REMYCGDGYRAVEITTHG-KNGMHVHVALLMPSGDVSENILIESGSDAMFDRWTSKLYSL 204
DB 140 RARYGVERTYSDYEVTDSWANGWHILRRMLLELRPLSDDELKAFEDSMFSRWSAGVKA 199
```



CC	acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
CC	hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
CC	dehydrogenases, terpene synthases, and cholesterol oxidase in an
CC	Actinomyetales bacteria. The replication protein or plasmid stability
CC	protein are useful in cloning and expression vectors and particularly in
CC	shuttle vectors for the expression of homologous and heterologous genes
CC	in Rhodococcus sp. This sequence represents the S. nigificans pSM22
CC	replication protein described in the disclosure of the invention.
CC	
XX	
XX	Sequence 451 AA;
XX	
XX	Query Match 16.1%; Score 322; DB 23; Length 451;
XX	Best Local Similarity 30.8%; Pred. No. 4.1e-23;
XX	Matches 101; Conservative 41; Mismatches 110; Indels 76; Gaps 15
QY	49 PISGV-----NGVTIVNGPKSGFGGILRSCGKMWICPCGAKYGAHRADISQVAHQI 102
DB	3 PASGVITAAQRAAGISVL-----GLMRCGRITWLCPCVCAATIRHRAEITAAVYEWI 54
QY	103 GTGSGVA-MVTMTMHTNAGORLHDIWTGISAAMKATNGRR-----WRT----- 145
DB	55 KRGGTAIVLVTFTAHGHDTDLADMDALQGTRKADAPRRPGAYORLITGWTAGRRRAKD 114
QY	146 -----REMYGCGGYRAVEITGK-NGMEVTHALLMFSGDV-----SENILESF 189
DB	115 GHRADBEGRIDRIGYVGMIRATEVTAVGQINGMPPHIAIYLVGGRTGERSAKQIVGTF 174
QY	190 --SDAMPDRMTSKIVSGFAAPLR-----NSGGIDVAKIGSEAD-OVIAAYLTATIA 237
DB	175 EPSRAALDEWGGQRAV-WTALAKVNPQFTPDDRHGVDPRKLETERRANDLAETIAQTQ 233
QY	238 SG--VGKEVSGSGDGKSGRHGRAPWEI---AVDAVG--DPQA-----LEIMREEFE 281
DB	234 DGKAPALEELARADIKTANGVAPFELLGRITGDLTGWTEDDAAGVGSLENNLARWHEYE 293
QY	282 FGSNGRAIAMSRLGRARAGIQAELTDA 309
DB	294 RATGGRRAIEMWTRYLRQMLGIDGGDTGA 321
XX	
XX	RESULT 8
XX	AA97537
XX	ID AA97537 standard; Protein; 427 AA.
XX	AC
XX	AA97537;
XX	DT 15-JAN-2001 (first entry)
XX	DE B. lactofermentum pSFK6 protein sequence.
XX	DE Temperature sensitive plasmid; TSCR; protein production;
XX	KW temperature sensitive replication control region; pSFK6.
XX	OS Brevibacterium lactofermentum.
XX	PN EPI038966-A1.
XX	PD 27-SEP-2000.
XX	PF 16-MAR-2000; 2000EP-0105326.
XX	PR 16-MAR-1999; 99JP-0069896.
XX	PA (AJIN ) AJINOMOTO CO INC.
XX	PI Nakamura J, Kanno S, Kimura E, Matsui K, Nakamatbu T;
XX	DR WPI; 2000-573832/54.
XX	DR N-PSDB; AAA90951.
PT	Plasmids containing a temperature sensitive replication control regions
PT	useful for breeding microorganisms for the production of amino acids by
PT	fermentation -

XX Disclosure; Page 19-22; 29pp; English.  
PS This sequence is the Brevibacterium lactofermentum pSF6 protein.  
XX The invention relates to a plasmid containing a temperature sensitive  
CC replication control region (TSRCR) and a marker gene (MG). The TSRCR is  
CC derived from plasmid pAM30 harboured by Brevibacterium lactofermentum  
CC (ATCC 13869) and allows the plasmid to replicate autonomously at a low  
CC temperature but does not allow the plasmid to replicate autonomously at  
CC an elevated temperature in coryneform bacteria within a temperature  
CC range in which the bacteria can grow. The plasmid can be used for  
CC modifying a chromosomal gene in a coryneform bacterium, which may be used  
CC for the production of useful substances, such as amino acids, by  
CC fermentation, to change their genetic traits. Therefore, the plasmid can  
CC be used for breeding microorganisms for the production of amino acids by  
CC fermentation. The plasmid comprises a TSRCR that allows the plasmid to  
CC replicate autonomously at an elevated temperature in coryneform  
CC bacteria within a temperature range in which the bacteria can grow.

XX Sequence 427 AA;

Query Match 15.9%; Score 318; DB 21; Length 427;  
Best Local Similarity 28.8%; Pred. No. 9.6e-23;  
Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

QY 34 LQOITSTETNACGR-----PISGVNVTITVNGPKSGGGLRSCKGMI CPCAAGV 86  
DB 27 MYKITSXALAGCHRRRDEAVAVSWSS-----NG--ASQFGLQNSHSRWGSSIALELV 79  
QY 87 GAHRADETISQVVAHOLGTGSVAVMTM--TMRHTAGORLHDLMTGSAAMKAATNGRRMTE 145  
DB 80 MGERRIELAIATKTNLAAGALMFMFVGVTRHRSQSFAQVEGITAAYSSVKTQMKKE 139  
QY 146 REMYGCDDGVYRAVEITHG-KNGMHHVHALMFGSDVSENILSESFDAMFDRMTSKLYSL 204  
DB 140 RARVGVHTYSYEVETDSMANGMHLHRNMLFLDRPLSDDELKAFEDSWFSRWSAGVYKA 199  
QY 205 GPAFLRNSG-GLD-VRKIGBADOVLAAYITKTASGVGMVSGSDGSGRRGNAPMEI 262  
DB 200 GMDAPLRHGVKLDQVSTWGDAAK-MATYIAK--GMSQELTGSATKTASKSYTPFGM 255  
QY 263 AVDAVGDPQALE-----LMREFEFGSMGRALAMSGRLARAGLGAELTDQI--- 311  
DB 256 -LDMTADSDAGEMDAVLVARMREYEVGSKNLRG-SMSRG--AKRALGIDYIDADVRE 311  
QY 312 VEOE-----ESAPVVAIIPARMMMTIRT 335  
DB 312 MEEELYKLAGLEAPERVESTRAVALVKKPDWKLIQS 348

RESULT 9  
AB06342  
ID AB06342 standard; Protein; 427 AA.

AC ABB06342;

DT 29-MAY-2002 (first entry)

XX Brevibacterium lactofermentum p48K protein sequence SEQ ID NO:8.

KM Kanamycin resistant gene; mutation; high temperature resistance;  
KW coryneform microbe; transformation.

OS Brevibacterium lactofermentum.

XX JP2002017362-A.

XX 22-JAN-2002.

XX 05-JUL-2000; 2000JP-0204236.

XX 05-JUL-2000; 2000JP-0204236.

XX

PA (AJIN ) AJINOMOTO KK.  
XX WPI; 2002-263243/31.  
DR N-PSDB; ABL49734.  
XX Transformation of high temperature-resistant coryneform microbe and  
PT transformant -  
PT Example; Page 22-23; 28pp; Japanese.

XX The present invention describes a method for transforming a high  
CC temperature-resistant coryneform microbe in which the high temperature-  
CC resistant coryneform microbe is treated with a chemical changing the  
CC structure of cell wall of the microbe and an electric pulse is applied  
CC to a solution containing the microbe having a charged structure of cell  
CC wall and a DNA. Also described are: (1) transforming a high temperature-  
CC resistant coryneform microbe in which the cell wall of a high  
CC temperature-resistant coryneform microbe is weakened and electric pulse  
CC is applied to a solution containing the microbe having weakened cell  
CC wall and a DNA; (2) a transformant of a high temperature-resistant  
CC coryneform microbe in which a recombinant DNA is introduced; and  
CC (3) a transformant of a high temperature-resistant coryneform microbe  
CC carrying both a plasmid vector derived from the high temperature-  
CC resistant coryneform microbe and a plasmid vector derived from  
CC coryneform glutamatum or Brevibacterium lactofermentum. The method  
CC is used for transforming a high temperature-resistant coryneform microbe.  
CC The present sequence represents a Brevibacterium lactofermentum p48K  
CC protein sequence, which is used in an example from the present  
CC invention.

XX Sequence 427 AA;

Query Match 15.9%; Score 318; DB 23; Length 427;  
Best Local Similarity 28.8%; Pred. No. 9.6e-23;  
Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

QY 34 LQOITSTETNACGR-----PISGVNVTITVNGPKSGGGLRSCKGMI CPCAAGV 86  
DB 27 MYKITSXALAGCHRRRDEAVAVSWSS-----NG--ASQFGLQNSHSRWGSSIALELV 79  
QY 87 GAHRADETISQVVAHOLGTGSVAVMTM--TMRHTAGORLHDLMTGSAAMKAATNGRRMTE 145  
DB 80 MGERRIELAIATKTNLAAGALMFMFVGVTRHRSQSFAQVEGITAAYSSVKTQMKKE 139  
QY 146 REMYGCDDGVYRAVEITHG-KNGMHHVHALMFGSDVSENILSESFDAMFDRMTSKLYSL 204  
DB 140 RARVGVHTYSYEVETDSMANGMHLHRNMLFLDRPLSDDELKAFEDSWFSRWSAGVYKA 199  
QY 205 GPAFLRNSG-GLD-VRKIGBADOVLAAYITKTASGVGMVSGSDGSGRRGNAPMEI 262  
DB 200 GMDAPLRHGVKLDQVSTWGDAAK-MATYIAK--GMSQELTGSATKTASKSYTPFGM 255  
QY 263 AVDAVGDPQALE-----LMREFEFGSMGRALAMSGRLARAGLGAELTDQI--- 311  
DB 256 -LDMTADSDAGEMDAVLVARMREYEVGSKNLRG-SMSRG--AKRALGIDYIDADVRE 311  
QY 312 VEOE-----ESAPVVAIIPARMMMTIRT 335  
DB 312 MEEELYKLAGLEAPERVESTRAVALVKKPDWKLIQS 348

RESULT 10  
AAE13531  
ID AAE13531 standard; Protein; 427 AA.

AC AAE13531;

DT 12-FEB-2002 (first entry)

XX B. lactofermentum p48K replication control region protein.

KM Coryneform bacteria; arginine repressor; L-arginine; liver function;  
KW amino acid infusion; mutant; mutain.



```
XX OS Brevibacterium lactofermentum.
XX EN EPI154020-A2.
XX PD 14-NOV-2001.
XX PF 24-APR-2001; 2001EP-0109457.
XX PR 28-APR-2000; 2000JP-0129167.
XX PA (AJIN ) AJINOMOTO KK.
XX PI Suga M, Asakura Y, Mori Y, Ito H, Kurahashi O;
XX DR WPI; 2002-012660/02.
XX DR N-PSDB; AAD22583.
XX PT Coryneform bacterium with inactivated arginine repressor, useful for
XX PT fermentative production of arginine at high levels -
XX PS Example 1; Page 20-21; 30pp; English.
XX CC The present invention relates to Coryneform bacterium in which the
XX CC arginine repressor does not function normally and which produces
XX CC L-arginine. Coryneform bacterium is used to produce L-arginine which is
XX CC useful in agents for promoting liver function and in amino acid infusions
XX CC or in 'comprehensive' amino acid pharmaceutical. Coryneform bacterium
XX CC accumulate L-arginine to higher levels than wild-type strains, e.g. 120
XX CC compared with 20, mg/dl. The present sequence is
XX CC Brevibacterium lactofermentum p48K used in the construction of shuttle
XX CC vector for Escherichia coli and Coryneform bacteria and temperature
XX CC sensitive vector.
XX SQ Sequence 427 AA;
Query Match 15.9%; Score 318; DB 23; Length 427;
Best Local Similarity 28.8%; Pred. No. 9.6e-23;
Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;
QY 34 LQGITTSERFNACGR-----PISGVNCTIVNGPKSGGPGGLRSCGKWCPCPCAGKY 86
DB 27 MKRTNISKALAGCHRRRDEAVAVSWSS-----NG--ASQFGLQNSHRSWSSLAELTV 79
QY 87 GAHRADEISQVVAHQLTGTSVAVMTM-TYRHTAGQRLHDLMTGLSAAMKAAATNGRRWTE 145
DB 80 MERRIELAIAIKNHLAAGALMRFYGTVRHNRSGSFAGVEAGIKTAYSMTKTSQMKTE 139
QY 146 REMYGGDGYVRAVEITHG-KNGMHVHVAHLLMFSGDVSENILSFSDAMFDRWTSKLYSL 204
DB 140 RARYGVHTYSDEYETDWSANQWHLRNMLLFLDRPLSDDELKAFEDSWFSRWSAGVVXA 199
QY 205 GAAPLRNSG-GLD-VRKIGEADQVLAAYLKIASGVGMEVSGSGKSGRGNAPMWI 262
DB 200 GNDAPLRHGVKLDQVSTWGDAAK-MATYTLAK--GMSQELTGSATKPAKSGSTYTFQM 255
QY 263 AYDAVAGDPOALE-----LMREFEFGSGRRATAIWSGRLRABAGLAELTDAQI--- 311
DB 256 -IDMLADQSDAGEDMDAVVAIWMREYEVSSKNLRS-SWGRG--AKRALGIDYIDADVRE 311
QY 312 VEOE-----ESAPYVVAIIPARSWMTIRT 335
DB 312 MEEELYKLAGLEAPERVESTRAVALVKPDWLTLOS 348
RESULT 11
ABB84281
ID ABB84281 standard; Protein; 456 AA.
XX ABB84281;
XX DT 13-JAN-2003 (first entry)
```

```
XX DE S. lividans pJ101 replication protein.
XX KW Plasmid stability protein; replication protein; ethylene forming enzyme;
XX KW carotenoid biosynthesis enzyme; isoprenoid; pyruvate decarboxylase;
XX KW polyhydroxyalkanoic acid synthase; P4A synthase; nitrile hydratase;
XX KW alcohol dehydrogenase; terpene synthase; cholesterol oxidase;
XX KW shuttle vector.
XX OS Streptomyces lividans.
XX PN WO200255709-A2.
XX PD 18-JUL-2002.
XX PF 12-DEC-2001; 2001WO-US47868.
XX PR 12-DEC-2000; 2000US-254868P.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PI Bramucci MG, Cheng Q, Kostichka KN, Tomb J;
XX DR WPI; 2002-557827/59.
XX PT New nucleic acid molecule encoding replication protein/plasmid
XX PT stability protein, useful in cloning and expression vectors,
XX PT particularly shuttle vectors for expression of heterologous genes in
XX PT Rhodococcus species -
XX PS Example 5; Fig 4A; 96pp; English.
XX CC This invention describes a novel nucleic acid encoding a replication
XX CC protein or a plasmid stability protein. The product of the invention is
XX CC useful for expression of nucleic acid such as genes encoding enzymes
XX CC involved in the production of isoprenoid molecules, polyhydroxyalkanoic
XX CC acid (PHA) synthases, carotenoid biosynthesis enzymes, nitrile
XX CC hydratases, ethylene forming enzyme, pyruvate decarboxylase, alcohol
XX CC dehydrogenase, terpene synthases, and cholesterol oxidase in an
XX CC actinomycetale bacteria. The replication protein or plasmid stability
XX CC protein are useful in cloning and expression vectors and particularly in
XX CC shuttle vectors for the expression of homologous and heterologous genes
XX CC in Rhodococcus sp. This sequence represents a Rhodococcus AN12 Rep
XX CC protein described in the disclosure of the invention.
XX SQ Sequence 456 AA;
Query Match 15.8%; Score 316; DB 23; Length 456;
Best Local Similarity 30.5%; Pred. No. 1.7e-22;
Matches 100; Conservative 40; Mismatches 112; Indels 76; Gaps 15;
QY 49 PISGV-----NGVTIVNGPKSGGPGGLRSCGKWCPCPCAGKVAHRADEISQVVAHQ 102
DB 3 PASGVIVAGTAAGTSVVL-----GLMRGRITLCEVCATINHKRAEITTAVEMVI 54
QY 103 GTGSVA-MVTMTRHTTAQQLHDLMTGLSAAMKAAATNGRR-----WRTE---- 145
DB 55 KKGGAAYIVTFFARHGHTDRDLADLALQGTREKTPDSRRPQAYORLTGTWAGRRAD 114
QY 146 -----REMYGGDGYVRAVEITHGK-NQMHVHVAHLLMFSGDV-----SENILSF 169
DB 115 GHRAADREGIRIRIGYVGMIRATEVTVAQINQMHHIAYLVGRTBGERGAKQIVATF 174
QY 190 --SDAMPRTWSKLYSLGFAAPLR-----NSGGLDVRKIGEAD-QVLAAYLTUKTA 237
DB 175 EPTGAALDEMGHMSV-WTALARKVNAFPTDDHGHVDPKLLETRDANDLAETIAKQ 233
QY 238 SG--VGMEVSGDGSGRGNAPWEI---AYDAVGG--DEQA-----LEIMREEE 281
DB 234 DQKAPALELARADLTATGAGVAVAPPELLGRIGDLTGWTEDDAAVGSILEMTLSRWHEY 293
QY 282 FSGMGRRAIWSRGRLRABAGLAELTDA 309
```

[illegible]







Db 312 MEEELYKLAGLEAPRVESTRAVALVXPDWKLIQS 348

RESULT 2

US-09-521-668B-20

Sequence 20, Application US/09521668B

Patent No. 6303383

GENERAL INFORMATION:

APPLICANT: NAKAMURA, JUN

APPLICANT: KANNO, SOHEI

APPLICANT: KIMURA, EICHIRO

APPLICANT: MATSUI, KAHIKO

APPLICANT: NAKAMATSU, TOSYOSHI

TITLE OF INVENTION: TEMPERATURE SENSITIVE PLASMID FOR CORYNEFORM BACTERIA

FILE REFERENCE: 0010-1093-0

CURRENT APPLICATION NUMBER: US/09/521,668B

CURRENT FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: JP 11-69896

PRIOR FILING DATE: 1999-03-16

NUMBER OF SEQ ID NOS: 20

SOFTWARE: PatentIn version 3.0

SEQ ID NO 20

LENGTH: 427

TYPE: PRT

ORGANISM: Brevibacterium lactofermentum

US-09-521-668B-20

Query Match 15.9%; Score 318; DB 4; Length 427;

Best Local Similarity 28.8%; Pred. No. 2,5e-24;

Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

QY 34 LQOITSETFNAAGR-----FISGVNGVITVNGPKSGFGGLRSCGKMTCPCCAGRY 86

Db 27 MKTITSKALAGCHRRRDEAVAVSWS-----NG--ASQFGLQSHSRMSSSLAELEV 79

QY 87 GAHRADISQVVAHQGTSSVAVMTM--TKRHTAGQTLHLMTGLSAAKAAINGRRKRE 145

Db 80 MGRRIELAIATGNHIAAGALMMPVGTVRHNSQSPAVQAGIKTAYSMTYSQMKKE 139

QY 146 REMYGGDGYRAVEITHG--KNGWHVHVHALLMPSGVSENILESFSDFAMFDRMTSLVSL 204

Db 140 RARYGVHTYSDEYVDSNANGWHLRNMLFLDRPLSDDELAFEDSMFSRWSAGVVA 199

QY 205 GFAPAPLRNSG-GID-VRKIGSEADQVLAAYLTIKIASGVMEVSSGDKSGRRGNRAPMEI 262

Db 200 GMDAPLRHGVKLDQVSTWGDAAK--MAYTLAK---GMSQELGSAITKASKSYPPFQM 255

QY 263 ANDAVGDDPOALE-----LMREFPFGSGRRALANSRGLRPARGLGALTDQI--- 311

Db 256 -IDMTLADQSDAGEDMDAVLVARWRREYEVSKNLR--SWSRG--AKRALGIDYIDAVRRR 311

QY 312 VEOE-----ESAPVVAIIPASMMWIRT 335

Db 312 MEEELYKLAGLEAPRVESTRAVALVXPDWKLIQS 348

RESULT 3

US-09-252-991A-17653

Sequence 17653, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 17653

LENGTH: 882

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

FEATURE:

NAME/KEY: UNSURE

LOCATION: (730)

OTHER INFORMATION: Identity of amino acid at the above locations are unknown.

US-09-252-991A-17653

Query Match 5.1%; Score 101.5; DB 4; Length 882;

Best Local Similarity 21.7%; Pred. No. 0.17;

Matches 87; Conservative 48; Mismatches 125; Indels 141; Gaps 20;

QY 47 GRPISGVNGVITVNGPKSGFGGLRSCGKMTCPCC--GKVGHRADISQVV----- 98

Db 487 GRVAGVGLALGQALGNR-----VAFVDAQGRKVGLEAELQSRPMHRPTE 535

QY 99 -----AQIGTGS-----VAMVTMTHTAGQRLHDL-----W 126

Db 536 IRHAPADGVGGAGRAVELGGGQQLQRLLVQVRIAL--HDQQRAGVGGRGRTGL 594

QY 127 TGLSAAMKAATNGRRMTEREMYGDDGYRA-----VEITH-----GKNGHV 169

Db 595 VGVAAGHGAVDQAAGRGDAPVLDPAAVVALAVLVASGHPVAFQVRLEVGCGGHA 654

QY 170 HV-----HALMPSGVSENILESFSDFAMFDRMTSLVSL-210

Db 655 GVGVAAVAGAEVDVHAL---AGDAGGVQAPAGAPVVLIGGVQAVEGLVADVLGLAAPT 711

QY 211 -----NNSGGLDVRKIGSEADQVLAAYLTIKIASGVMEVSSGDKSGRRGNRAPM 260

Db 712 VDRPFGVQRLVGGLEIKRVRRAEQEAVVL-----VVGDVHLCRVGH----- 756

QY 261 EIAVDVAG-GDPOALEM-----REFPFGSGRRALANSRGLR-ARAG 301

Db 757 PVHADAVARGAAGDADVAVGVIVAHRAADERRAVIDGAAGRRLVABGGGIMARVE 816

QY 302 LGAEITDQIYQESAPVMAIIPASMMWIRTCAPIYFG 342

Db 817 TGVHLADLHALADAGIGLVGPHFPQAVLL-----VFG 851

RESULT 4

US-09-252-991A-30345

Sequence 30345, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 30345

LENGTH: 1395

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30345

Query Match 5.0%; Score 101; DB 4; Length 1395;

Best Local Similarity 24.5%; Pred. No. 0.37;

Matches 104; Conservative 34; Mismatches 162; Indels 124; Gaps 24;

QY 13 DRPVLVSSDKGIRHELRPKLQOITSETFNAAGR--PISG-----VNGVITVNGPKSSG 66

Db 511 DRIAVAVGDBHRAAR--FSAAEAAVAVQDGRGRRAIDGGAALAAAGVARRIGRGVD 568

QY 67 FGLRS-CGKMTCPCCAGKVAHRADISQVV-----AHQGL-----TOSVAMVTMTMHH 116

```

Db      569 HOSTIQRKRG- EAVVAAA-VGGDLDVAVAVAGGHAARLGAALGAERGAAGVD- -H 622
QY      117 TAGQ--RLHDLMTGLSAAMKAATNG-----RRMTEREMTGC-----DGYV 155
Db      623 WRGRGVAVH---GDTAARTGAVAGVGGRVDHAGVGRRGSEAPVAALIGDLPDGA 678
QY      156 RAVEITTECKNGWHVHVHALLMFSGDVSENILESFDAMFDRMTSKLVSLGFAPLR- - 211
Db      679 VAIGQGHG-----GARLGAALERGAVARVDHRRGRCHGVHGDAGRTGVA 723
QY      212 ---NGGLDVYKIG-----GEADQVLAATLTIASGVCMENVSQGGKSGRHNRAPEL 263
Db      724 GRVGRGVDRHVRVGGRRRAGE- PVAAGVGDLANRVAVAVQGGHG-SARLGAALERTV 781
QY      264 VDAVAGDPOALELMREFEFGSGMGR-ALAW-----SRGLR- - -ARAGLGA 305
Db      782 ARIDGRGR-----FDGIGNAGKRTAVAGGIPGHVHDHGTIRQGRREVRVAVIGGG 836
QY      306 LTDAQIVGEESAPYVVALIPARSMWMTITCAPYVFGELIGLVEAGATWENLDHLYRL 365
Db      837 LAD-----RIMAVGQGHAAARFGSPAPFAVARL-----DHRVYR 873
QY      366 PAAD 369
Db      874 HAVD 877

```

```

RESULT 5
US-09-252-991A-23181
; Sequence 23181, Application US/09252991A
; Patent No. 6551795

```

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23181
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23181

```

```

Query Match      5.0%; Score 99.5; DB 4; Length 676;
Best Local Similarity 25.6%; Pred. No. 0.18;
Matches 89; Conservative 31; Mismatches 135; Indels 93; Gaps 19;

```

```

QY      62 PRGSGFGILRSCGKMTIC--PCCAGKVAHRADIS-----QVVAHQLTGSSVAMVTMT 113
Db      318 PRGPAAGRRASRGCGCGCRGCGAGRTAGRASPVATATRGSCSRAPGAPVA----- 372
QY      114 MHTTAGQRLHDLMTGLSAAMKAATNRKRTEREMTGCQGYRAVEITTECKNGWHVHNTA 173
Db      373 -GRGANRRRRSRWSLPAS--CLSRGR-----GYCSVARRRPG- - - 409
QY      174 LMFSGDVSENILESFDAMFDRMTSKLVSLGFAPLRNSGGLDVKIGEADQVLAATL 233
Db      410 -AVRGSVSCALLTQAKRAGD-----AVDLGIGA-----YLLDNRFBENARQFLAAGR 458
QY      234 TKIAGVG-MEVGS-----GDGK-----SGRHGNRAPEIADV-AVGD-POALELME 279
Db      459 GRRRAAFETEGALFFQAHGEGRLDVRQAGPAQFGFVADHQAAGDVAAQALE----- 514
QY      280 FFGSGNR--RAIAMSRRGLARAGLGAELTDAQIVGEESAPV--MVAIIPARSMWMTIR 334
Db      515 -----GRWVAQTEQTEGVAE-----GQDVGEAQVAAQRRGRPHRSAAVVVDDAQSHHLLA 564

```

```

QY      335 TCAPYV-----FGELIGVEAGA--TWENLRDHLHYRLPAAD 369
Db      565 GRAALVEHLEEDHVPPRHDEGQQQAAGAGELAEERQAEHQHQDQAAD 612

```

```

RESULT 6
US-09-252-991A-19231
; Sequence 19231, Application US/09252991A
; Patent No. 6551795

```

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19231
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19231

```

```

Query Match      4.9%; Score 98.5; DB 4; Length 621;
Best Local Similarity 22.4%; Pred. No. 0.21;
Matches 62; Conservative 27; Mismatches 87; Indels 101; Gaps 11;

```

```

QY      175 LMFSGDVSENILESFDAMFDRW-----TSKLVSLGFAPLRNSGGL 216
Db      323 LTVSGDVSNLIGSLTDLPQLRPHLCLEDESLLPHRLQFELRQEFMLRRAGV- 381
QY      217 DVKRTGSEADQVLAATLTIASGVCM-----VSGDGKSGRHNRAPEIADVAGDQ 272
Db      382 ---NFGATRMSSLLMYQLADIGEGCYHCVTLAGE-----ADELFWGYR 425
QY      273 ALEMRPEFGSMGRATAMSGRLARAGLGAELTDAQIVG-ESSA----- 318
Db      426 HLELMRRDAPRRPFAAAMEYRRKALAEPRGRVARETELEHAEALQGLEAIG 485
QY      319 --PVVAIIP-----ARSMWMTITC--APYVG-----ETLG-----LVE 349
Db      486 QEDLHYSLPELRRADHLLMSRTIARPYLHGALAQARGLRVGDPAKAPVALLEQ 545
QY      350 AGATWE-----NURDHLHYRLPA 367
Db      546 REKRWQAQPKRHFRLPFERWPQALGEMRHLAERLPA 582

```

```

RESULT 7
US-09-252-991A-22282
; Sequence 22282, Application US/09252991A
; Patent No. 6551795

```

```

; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22282
; LENGTH: 1409
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

```



US-09-252-991A-22282

Query Match 4.9%; Score 98.5; DB 4; Length 1409;  
Best Local Similarity 23.8%; Pred. No. 0.69;  
Matches 103; Conservative 51; Mismatches 160; Indels 119; Gaps 24;

QY 14 RPPVLSVRKGRIRHELR--PKIQITTSSTFNACGRPIISGVNGVTI-----VNGP 62  
DB 275 RPPALDQASGGILQVORRCFGLQLSR---RLASRLDAGPAGETVRPRPHGLILCY 330  
QY 63 KSGS-----FCGRSGCGKMIC--PCCAG---KUGAHRADEISQVVAHQ 101  
DB 331 RQSGATDHPVAGHALLLEIFGQIQAIPEVVAADPHLALQORAGABAETITQCAQIQ 390  
QY 102 L-----GT---GSVAMVTMTMRTAGCR--LHDLMTGLSAWKATN----- 138  
DB 391 LAIAGLRQGTGQGVAAATAQSLAVGQRRGRGRIEQTALISQAVAAAGNPSAVAV 450  
QY 139 GRP---WRTEREMYGCDGYRAVEITHGKGMHVHALLMPSGDVSENILSFSFDMED 195  
DB 451 GQRIVTLRQOQVAPQAGIOPVAVAHQDA-----VPAAGIDETI-----DRPVR 496  
QY 196 RMTSKVSLGFAPLNSGGLDVRKIGBADQVLAAYTKIASGVMEVSGDGKSGRHG 255  
DB 497 RLRQHVAGLQAAAMGADVVD--RHVVGSD--CLAI-----GQQRADVADALAGRH 548  
QY 256 NPAPWEIADVAVGDPQ--ALELMREFEFGSMGRRAIAMSRLRARAGIGALTDQIV 312  
DB 549 QAAQLE-AVEQFVAGQAAALALGHSVCGRLGLRLAGTDHLLAAGIGAE---SVL 603  
QY 313 EDEESAPVVAIIP-----ASMMIRTCAPVFEIILGLVBAQTWB-----NLR 358  
DB 604 QQAALAHVVGVAHPDQLCAARRGIL-----RLVVAVQAG-LMRFGLPLAAGVA 652  
QY 359 DHL--HYRLPAD 369  
DB 653 DHLAVEPRIGAAD 665

## RESULT 8

US-09-252-991A-30935  
; Sequence 30935; Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30935  
; LENGTH: 1245  
; TYPE: PR  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30935

Query Match 4.9%; Score 97.5; DB 4; Length 1245;  
Best Local Similarity 21.2%; Pred. No. 0.73;  
Matches 86; Conservative 50; Mismatches 143; Indels 127; Gaps 19;

QY 9 LSGKRP-PYLVSDKRGIRHELRPKIQITTSSTFNACGRPI-----SGVNGVT 57  
DB 188 LRQPRSGPAPFGRRQRPGRHRTYPVQ-----RPAQAPRRGGGTGARVA 235  
QY 58 IVNGPKSGFGGRLSGCGKGM-----ICPCAGKVGARADEISQVVAHQ 102  
DB 236 AVEQE-----GCRITCAAGRRQAPRPRRAAPGTRPCGAGAGRLAWATGVRQAAGRT 269

QY 103 GTGSVAMVTMT-----MRHTAGORLHDL-----WTGLSAWKATNCR 140  
DB 290 GGGVLAISRRAADPPGRSPACQAGIRYGADRPFRFSPRPAAGATAGAAAGW----- 342  
QY 141 RWRTEREMYGCD-GYRAVEITHGKGMHVHALL-----MFGSDVSENILSFS 190  
DB 343 -WR-----GADPGALRRV-VSINENIFSHMTWFLPASLSLVAAGDIQQSIGESS 394  
QY 191 DAM-FDR--WTSKVLISGFAPLNSGGLDVRKIGBADQVLAAYTKIASGVMEVSG 247  
DB 395 MTLPTTRAAMPPLCSAAVLGAALMWAASAAERRFDLPQPLAASLSRLAQOQVYFLD 454  
QY 248 DKSGRHGNRAPWEIADVAVGDPQALE--LMREFEFGSMGRRAIAMSRLRARAGIGA 304  
DB 455 R--SLRLGIRAP--ALSQSYGVREALRLVGSSELVEAGDGYVARRQVDAYSNAL 509  
QY 305 ELTDQIV-----EDESAPVVAIIP 326  
DB 510 QL-DAQITVGNRGREVDASNGRSTLTRDIERQADNIPSLIQITP 554

## RESULT 9

US-09-335-409-11  
; Sequence 11; Application US/09335409  
; Patent No. 6121029  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goettlich, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/335,409  
; CURRENT FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 713  
; TYPE: PR  
; ORGANISM: Sorangium cellulosum  
US-09-335-409-11

Query Match 4.8%; Score 97; DB 3; Length 713;  
Best Local Similarity 22.1%; Pred. No. 0.36;  
Matches 78; Conservative 38; Mismatches 157; Indels 80; Gaps 12;

QY 30 LRPKIQITTSSTFNACGRPIISGVNGVTIYNGPKSGFGGRLSGCGKMICPCAGVGAH 89  
DB 302 MRVDVSQLTPAPAMGTVALLLATATPAKVPALGRLGLSGSEALVAVGLNMG-- 359  
QY 90 RADEISQVVAHQIGTGS-----VAMVTMTMRTAGQRLHDLMTGLSAWKATNCR 140  
DB 360 -TDLVAIVGVELGILSNAYTMVAVALTV-----TABALL-ITLHKRA---PPTQSE 410  
QY 141 RWRTEREMYGCDGYRAVEITHGKGMHVHALLMPSGDVSENILSFSFDMEDWTSK 200  
DB 411 SARLEREEARARAYIPGE---RLVPIVAHALPEFANDIVESIVAS----- 454  
QY 201 LVSIGFAPLNSGGLDVRKIGBADQVLAAYTKIASGVMEVSGDGKSGRHNR--- 257  
DB 455 -----KRLGFTVD-ITELSVQQAQPPSRAAGASRLAAGLARGLRV 496  
QY 258 APWEIADVAVGDPQALELMREFEFGSMGRRAIAMSRLRARAGIGALTDQIVQES 317  
DB 497 GTRQRRRLRGSTQALIRASRDHDLVIGARSPARAKGSF-----GRLODAIVQRAESN 551  
QY 318 APVVAIIPASMMIRTCAPVFGIILGLVEAGATWENIRDLHTRLPADV 370  
DB 552 VLVVGDPPAAE---RASHRIIVPIIG-----LEVSPAADL 586

```

RESULT 11
US-09-567-969-11
; Sequence 11, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goeschl, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIONONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0

```

[illegible]

Db 411 SARLEREEAARRAYIPGVE-----RILVPIVAHALPGFATDIVESIVAS----- 454  
Qy 201 LVSLGFAPLRNNSGLDVKRIKIGEADQVLAAYLTIKIAGVGMVEVSGDGKSGRHGNR--- 257  
Db 455 -----KRIKIGETVD-ITELSVGOAPGPPSRAAGEASRGARLGARLRV 496  
Qy 258 APWEIADVAVGDPOALELMREFEFGSMGRRAIAMSRLRAGIGAEITDQIIVEQSES 317  
Db 497 GIWRQRRELRSIGAILRASRDHDLVIGARSPARAKGMSF-----GRLODAIVORASBN 551  
Qy 318 APVWVAIIPARSMWIRTCAPYVFGELIGLVEAGATWENLDHLYRLPADV 370  
Db 552 VLVVVGDPPEAAE---RASARRILVPIIG-----LEYSFAADL 586

RESULT 13  
US-09-568-486-11  
; Sequence 11, Application US/09568486  
; Patent No. 6355459  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goelach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,486  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 713  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-486-11

Query Match 4.8%; Score 97; DB 4; Length 713;  
Best Local Similarity 22.1%; Pred. No. 0.36;  
Matches 78; Conservative 38; Mismatches 157; Indels 80; Gaps 12;  
Qy 30 LRPKLOQITSETFNACGRPIGNGVGTIVNGPKSGFGGLRSCGKWCPCCAKVGAH 89  
Db 302 MRVDVSQLRTPAAGVVALLLATATPAKVPALGRLGGLGSEBALVAVGLNMKG-- 359  
Qy 90 RADEISQVVAHQLTGTS-----VAVVTWTRHTAGORLHDLTGLSAAMKAATNGR 140  
Db 360 -TDLVAIVGVELGILSNAYTMVAVALTV-----TASPALL-ITLEKRA---PPTQEE 410  
Qy 141 RWRTEREMYGCDGYRAVEITHGKNGWHVHALMFSGDVSENILSFSDAMPFRWTSK 200  
Db 411 SARLEREEAARRAYIPGVE-----RILVPIVAHALPGFATDIVESIVAS----- 454  
Qy 201 LVSLGFAPLRNNSGLDVKRIKIGEADQVLAAYLTIKIAGVGMVEVSGDGKSGRHGNR--- 257  
Db 455 -----KRIKIGETVD-ITELSVGOAPGPPSRAAGEASRGARLGARLRV 496  
Qy 258 APWEIADVAVGDPOALELMREFEFGSMGRRAIAMSRLRAGIGAEITDQIIVEQSES 317  
Db 497 GIWRQRRELRSIGAILRASRDHDLVIGARSPARAKGMSF-----GRLODAIVORASBN 551  
Qy 318 APVWVAIIPARSMWIRTCAPYVFGELIGLVEAGATWENLDHLYRLPADV 370  
Db 552 VLVVVGDPPEAAE---RASARRILVPIIG-----LEYSFAADL 586

RESULT 14  
US-09-568-472-11  
; Sequence 11, Application US/09568472

; Patent No. 6358719  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goelach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/568,472  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 713  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-568-472-11

Query Match 4.8%; Score 97; DB 4; Length 713;  
Best Local Similarity 22.1%; Pred. No. 0.36;  
Matches 78; Conservative 38; Mismatches 157; Indels 80; Gaps 12;  
Qy 30 LRPKLOQITSETFNACGRPIGNGVGTIVNGPKSGFGGLRSCGKWCPCCAKVGAH 89  
Db 302 MRVDVSQLRTPAAGVVALLLATATPAKVPALGRLGGLGSEBALVAVGLNMKG-- 359  
Qy 90 RADEISQVVAHQLTGTS-----VAVVTWTRHTAGORLHDLTGLSAAMKAATNGR 140  
Db 360 -TDLVAIVGVELGILSNAYTMVAVALTV-----TASPALL-ITLEKRA---PPTQEE 410  
Qy 141 RWRTEREMYGCDGYRAVEITHGKNGWHVHALMFSGDVSENILSFSDAMPFRWTSK 200  
Db 411 SARLEREEAARRAYIPGVE-----RILVPIVAHALPGFATDIVESIVAS----- 454  
Qy 201 LVSLGFAPLRNNSGLDVKRIKIGEADQVLAAYLTIKIAGVGMVEVSGDGKSGRHGNR--- 257  
Db 455 -----KRIKIGETVD-ITELSVGOAPGPPSRAAGEASRGARLGARLRV 496  
Qy 258 APWEIADVAVGDPOALELMREFEFGSMGRRAIAMSRLRAGIGAEITDQIIVEQSES 317  
Db 497 GIWRQRRELRSIGAILRASRDHDLVIGARSPARAKGMSF-----GRLODAIVORASBN 551  
Qy 318 APVWVAIIPARSMWIRTCAPYVFGELIGLVEAGATWENLDHLYRLPADV 370  
Db 552 VLVVVGDPPEAAE---RASARRILVPIIG-----LEYSFAADL 586

RESULT 15  
US-09-567-899-11  
; Sequence 11, Application US/09567899  
; Patent No. 6383787  
; GENERAL INFORMATION:  
; APPLICANT: Schupp, Thomas  
; APPLICANT: Ligon, James  
; APPLICANT: Molnar, Istvan  
; APPLICANT: Zirkle, Ross  
; APPLICANT: Cyr, Devon  
; APPLICANT: Goelach, Joern  
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
; FILE REFERENCE: 4-30582A  
; CURRENT APPLICATION NUMBER: US/09/567,899  
; CURRENT FILING DATE: 2000-05-10  
; PRIOR APPLICATION NUMBER: 09/335,409  
; PRIOR FILING DATE: 1999-06-17  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 713  
; TYPE: PRT



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 23:02:33 ; Search time 232 Seconds  
(without alignments)  
280.573 Million cell updates/sec

Title: US-10-007-527A-2

Perfect score: 2005  
Sequence: 1 MTSVSAEHLSGKDRPPVLVS.....HLHTRLPADYVPPIISVRK 379

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters: 644079

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCIT5_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2005	100.0	379	US-10-007-527A-2	Sequence 2, Appli
2	2005	100.0	379	US-10-007-452-2	Sequence 2, Appli
3	609.5	30.4	459	US-10-007-527A-21	Sequence 21, Appli
4	609.5	30.4	459	US-10-007-452-21	Sequence 21, Appli
5	326	16.3	427	US-09-835-381-6	Sequence 6, Appli
6	322	16.1	451	US-10-007-527A-24	Sequence 24, Appli
7	322	16.1	451	US-10-007-452-24	Sequence 24, Appli
8	318	15.9	427	US-09-835-381-8	Sequence 8, Appli
9	318	15.9	427	US-10-196-232-7	Sequence 7, Appli
10	316	15.8	456	US-10-007-527A-22	Sequence 22, Appli
11	316	15.8	456	US-10-007-452-22	Sequence 22, Appli
12	282	14.1	466	US-09-826-191-9	Sequence 9, Appli
13	282	14.1	466	US-10-263-666-9	Sequence 9, Appli
14	282	14.1	466	US-10-261-481-9	Sequence 9, Appli
15	282	14.1	466	US-10-261-942-9	Sequence 9, Appli

16	266.5	13.3	528	US-10-007-527A-23	Sequence 23, Appli
17	266.5	13.3	528	US-10-007-452-23	Sequence 23, Appli
18	103.5	5.2	417	US-10-156-761-10798	Sequence 10798, A
19	103	5.1	647	US-10-021-660-94	Sequence 94, Appli
20	100.5	5.0	544	US-10-156-761-8540	Sequence 8540, Ap
21	97	4.8	713	US-10-014-717-11	Sequence 11, Appli
22	96	4.8	7746	US-10-156-761-7965	Sequence 7965, Ap
23	95.5	4.8	3798	US-10-014-717-6	Sequence 6, Appli
24	95.5	4.8	4471	US-10-205-032-10	Sequence 10, Appli
25	94.5	4.7	877	US-10-156-761-13758	Sequence 13758, A
26	94	4.7	6146	US-10-156-761-10436	Sequence 10436, A
27	93.5	4.7	356	US-09-764-853-570	Sequence 570, App
28	92	4.6	310	US-10-156-761-9467	Sequence 9467, Ap
29	91.5	4.6	981	US-10-156-761-15071	Sequence 15071, A
30	91.5	4.6	5245	US-10-329-079-11	Sequence 11, Appli
31	91	4.5	245	US-10-156-761-14296	Sequence 14296, A
32	91	4.5	956	US-09-815-242-11925	Sequence 11925, A
33	90.5	4.5	309	US-10-306-762-88	Sequence 88, Appli
34	89.5	4.5	585	US-09-728-626-6032	Sequence 6032, Ap
35	89.5	4.5	3352	US-10-156-761-7961	Sequence 7961, Ap
36	89	4.4	3739	US-09-861-289-33	Sequence 33, Appli
37	89	4.4	3739	US-09-860-846-33	Sequence 33, Appli
38	89	4.4	3739	US-09-988-3848-33	Sequence 33, Appli
39	89	4.4	3739	US-09-793-708-2	Sequence 33, Appli
40	89	4.4	3739	US-10-201-365-3	Sequence 2, Appli
41	89	4.4	3739	US-10-160-539-2	Sequence 3, Appli
42	89	4.4	3739	US-10-271-889-33	Sequence 2, Appli
43	89	4.4	6291	US-10-329-079-41	Sequence 41, Appli
44	89	4.4	6291	US-10-329-079-41	Sequence 41, Appli
45	89	4.4	11877	US-09-861-289-6	Sequence 6, Appli

## ALIGNMENTS

```
RESULT 1
US-10-007-527A-2
; Sequence 2, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bremucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Koestelchka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: C11709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; CURRENT FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Rhodococcus AN12
US-10-007-527A-2
```

Query Match	Score	DB	Length	Indels	Gaps
Best Local Similarity	100.0%	Pred. No. 5.3e-195			
Matches	379	Conservative	0	Mismatches	0
QY	1	MTSVSAEHLSGKDRPPVLVS	60		
DB	1	MTSVSAEHLSGDRPPVLVS	60		
QY	61	GPKSGGFGALRSCGKGWICPCGAGKVGARHRADEISQVVAHQIGTGSVAMVTMTHTAGQ	120		
DB	61	GPKSGGFGALRSCGKGWICPCGAGKVGARHRADEISQVVAHQIGTGSVAMVTMTHTAGQ	120		
QY	121	RLHDMTGLSAAWKATNGRWRTERENYGCQGYRAYEITGKXGKHVHVALMESGD	180		
DB	121	RLHDMTGLSAAWKATNGRWRTERENYGCQGYRAYEITGKXGKHVHVALMESGD	180		

```

Qy 181 VSENILSFSDAMPDRMTSKVLISGFAPLRNSGGLDVRIKIGEGADQVLAAYLTXTIASGV 240
Db 181 VSENILSFSDAMPDRMTSKVLISGFAPLRNSGGLDVRIKIGEGADQVLAAYLTXTIASGV 240
Qy 241 GMEVSGDGKSGRHNRAPEIADVAGDPOALELMEFEFGSGRRALAMSRGLRARA 300
Db 241 GMEVSGDGKSGRHNRAPEIADVAGDPOALELMEFEFGSGRRALAMSRGLRARA 300
Qy 301 GIGAEITDAQIVQEESAPVMAIIPARSMWMTIRTCAPYVFGELIGLVEAGATWENLRDH 360
Db 301 GIGAEITDAQIVQEESAPVMAIIPARSMWMTIRTCAPYVFGELIGLVEAGATWENLRDH 360
Qy 361 LHYRLPADVPRPPIISVRK 379
Db 361 LHYRLPADVPRPPIISVRK 379

```

```

RESULT 2
US-10-007-452-2
; Sequence 2, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 379
; TYPE: PRF
; ORGANISM: Rhodococcus AN12
US-10-007-452-2

```

```

Query Match 100.0%; Score 2005; DB 15; Length 379;
Best Local Similarity 100.0%; Pred. No. 5,3e-195;
Matches 379; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSVSAEHLISGKDRPVVYSSDKRGIRHRLRPKLOITTSFTFNACGRPISGVNGVTYN 60
Db 1 MTSVSAEHLISGKDRPVVYSSDKRGIRHRLRPKLOITTSFTFNACGRPISGVNGVTYN 60
Qy 61 GPKSGFGGIRSGCGKMTCPCCAGKVGARADEISQVVAHQLGTSVAVMTMTMTHTAQ 120
Db 61 GPKSGFGGIRSGCGKMTCPCCAGKVGARADEISQVVAHQLGTSVAVMTMTMTHTAQ 120
Qy 121 RLHDLMTGLSAANKAATNGRRWTEREMTGCDGYVAVEITHGKNGMHVHVALLMFSGD 180
Db 121 RLHDLMTGLSAANKAATNGRRWTEREMTGCDGYVAVEITHGKNGMHVHVALLMFSGD 180
Qy 181 VSENILSFSDAMPDRMTSKVLISGFAPLRNSGGLDVRIKIGEGADQVLAAYLTXTIASGV 240
Db 181 VSENILSFSDAMPDRMTSKVLISGFAPLRNSGGLDVRIKIGEGADQVLAAYLTXTIASGV 240
Qy 241 GMEVSGDGKSGRHNRAPEIADVAGDPOALELMEFEFGSGRRALAMSRGLRARA 300
Db 241 GMEVSGDGKSGRHNRAPEIADVAGDPOALELMEFEFGSGRRALAMSRGLRARA 300
Qy 301 GIGAEITDAQIVQEESAPVMAIIPARSMWMTIRTCAPYVFGELIGLVEAGATWENLRDH 360
Db 301 GIGAEITDAQIVQEESAPVMAIIPARSMWMTIRTCAPYVFGELIGLVEAGATWENLRDH 360
Qy 361 LHYRLPADVPRPPIISVRK 379
Db 361 LHYRLPADVPRPPIISVRK 379

```

```

RESULT 3
US-10-007-527A-21
; Sequence 21, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 459
; TYPE: PRF
; ORGANISM: Arcanobacterium pyogenes
US-10-007-527A-21

```

```

Query Match 30.4%; Score 609.5; DB 15; Length 459;
Best Local Similarity 36.3%; Pred. No. 3.8e-53;
Matches 148; Conservative 62; Mismatches 133; Indels 65; Gaps 15;

Qy 5 SAEHLISGKDRPVVYSSD-----KGIHRLRPKLOITTSFTFNACGR-PTSGVNGV 56
Db 67 SATHPGLGTVLTPVSNKSKTKAKSRKERYELDLGLISTISVRKCGVPAVPLVSL 126
Qy 57 TVVNGPKSGFGGIRSGCGKMTCPCCAGKVGARADEISQVVAHQLGTSVAVMTMTMR 115
Db 127 RAKSDGKAGVGGHGTGSGWACPVCSAKIABRTDQVVDHAVKMGMTVSMULTQR 186
Qy 116 HTAGRLHDLMTGLSAANKAATNGRRWTEREMTGCDGYVAVEITHGKNGMHVHVALL 175
Db 187 HHKGQGLHMDALSTAMNRVYSGRRWTEFKEQFGLVGVYANETTHGKHGHHVSHYLI 246
Qy 176 MESGDVSENILSF-----SDAMPDRMTSKVLISGFAPLRNSGG 215
Db 247 ISEKD--PLVSTFYQKQKRRRLPYPEIYMSDFTAEKWEAGLAHGVDF-LRDSGG 302
Qy 216 L-----DVRIKIGEGADQVLAAYLTKI--ASGVMEVSGDGKSGRHNRAPEIADV 266
Db 303 LDMYAKDAPRAGN-----YISKMGTSIDAISSEVTLGSKFKAARNGRTPPQILADI 354
Qy 267 VG-GDPQALELMEFEFGSGRRALAMSRGLRARAAGLTDQIVQEESAPVMAIIP 325
Db 355 LSLGVDVDDLKLMKEYEKASFGRRALTWKGLRDMANLGVGEQDEDEIA-SEETGDEAIALF 413
Qy 326 PARSMWMTIRTCAPYVFG--EILGLVEAGA-----TWENLRDHRLHRLP 366
Db 414 THDAWRQYR-----FGAEILDVYESGRRAAAYRWLPFR--IDWSLP 455

```

```

RESULT 4
US-10-007-452-21
; Sequence 21, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; NUMBER OF SEQ ID NOS: 30

```

SOFTWARE: Microsoft Office 97  
SEQ ID NO 21  
LENGTH: 459  
TYPE: PRF  
ORGANISM: Arcanobacterium pyogenes  
US-10-007-452-21

Query Match 30.4%; Score 609.5; DB 15; Length 459;  
Best Local Similarity 36.3%; Pred. No. 3.8e-53;  
Matches 148; Conservative 62; Mismatches 133; Indels 65; Gaps 15;

QY 5 SAEHLSGKDRPPVLYSSD-----KRGIRHRLRKLQOITTSFPMNAGR-PISGVAGV 56  
DB 67 SATHPAGVTVLTFFPVSNESEKTKTAKRSRSEYELRGLEAISTIESVRKCGRPVAPLVSL 126  
QY 57 TVNPPKSGSGFGLSCGKMWICPCGAGVGAHRADEISQVVAHOLGTG-SVAMVTMTMR 115  
DB 127 RAKSGKAGVGGTLTCCSWACPVCSAKIAPRKTDLQGVDAVAKGMTSMITLQK 186  
QY 116 HTAGORLHDLMTGLSAAMKAAATNGRRMTEREMYGCDGVRAVBITTGKNGMHVHALL 175  
DB 187 HHKGGGLKHLMDALSTAMNRVTSGRRWTEFEKGFGVGVYRANEITHGKGMHVASHVLI 246  
QY 176 MFGSGVSENIIESF-----SDAMPRTSTSLVSLGPAAPLRNSGG 215  
DB 247 ISEKD---PLTSTFYORKQGRRLPYPEIYMSDPIAFERWEAGLAKGVDF-LRDSGG 302  
QY 216 L-----DYRKIGGRADQVLAAYLTKI--ASGVGMEVSGDCKSGRGNRAPMELAVDA 266  
DB 303 LDMVYAKDKRAIAGN-----YVSKQCTSTAISSVTLGGPKAKRNGNRITPFQLLADI 354  
QY 267 VG-GDPQALMLWREFEFGSGMRRALAMSRGLRARAAGLGAELTDAQIVEOESAPVWALI 325  
DB 355 LSLGVDVDLKMLKEYEKASFGRRLTWSKGLDMANLVGEQDEILA-SEIIGDEAIALF 413  
QY 326 PARSMNTITTCAPYVFG--EITGLVEAGA-----TWENIRDLIHLRLP 366  
DB 414 THDARQYVR-----FGAAELLDVYESGGRAAAYRWLDFRE-IDWSLP 455

RESULT 5  
US-09-835-381-6  
Sequence 6, Application US/09835381  
Patent No. US20020045223A1  
GENERAL INFORMATION:  
APPLICANT: SUGA, Mikiko  
APPLICANT: ASAKURA, Yoko  
APPLICANT: MORI, Yukiko  
APPLICANT: KURAHASHI, Osamu  
TITLE OF INVENTION: ARGININE REPRESSOR DEFICIENT STRAIN OF CORYNEFORM BACTERIUM AND N  
FILE REFERENCE: 206018USO  
CURRENT APPLICATION NUMBER: US/09/835,381  
CURRENT FILING DATE: 2001-04-17  
PRIOR APPLICATION NUMBER: JP2000-129167  
PRIOR FILING DATE: 2000-04-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 6  
LENGTH: 427  
TYPE: PRF  
ORGANISM: Brevibacterium lactofermentum  
US-09-835-381-6

Query Match 16.3%; Score 326; DB 9; Length 427;  
Best Local Similarity 29.1%; Pred. No. 2.2e-24;  
Matches 98; Conservative 58; Mismatches 131; Indels 50; Gaps 15;  
QY 34 LQOITTSFPMNAGR-----PISGVNVTIYNGKSGFGRLSCGKMWICPCGAGV 86  
DB 27 MYKITSKALAGCHWRDEAVAVSWSS-----NG--ASQFGLQNSHRSWSPILAELV 79

QY 87 GAHRADEISQVVAHOLGTGSVAMVTM--TMRHHTAGORLHDLMTGLSAAMKAAATNGRRMTE 145  
DB 80 MEERREIELAIYTKHLLAAGALMFMVGVVRNRSQSFAQVAGIKTAISSVWKTQSQWKE 139  
QY 146 REMYCDGYVAVEITHG-KNGMHVHVALIMFSGSENILIESFSDDAMPRTSTSLVSL 204  
DB 140 PARYGVHTTSDYVETDSKNGMHLHRMLLFLRPLSLDDDLKAFEDSMFSRMSAGVYKA 199  
QY 205 GPAAPLRNSG-GLD-VRKIGGEADQVLAAYLTJKIASGVGMEVSGDCKSGRGNRAPMEI 262  
DB 200 GMDAPLRHGVKLDQVSTWVGDAK--MATYLAK---GMSQELTGSATKTSKGSYTPQM 255  
QY 263 AYDAVGDPQALF-----LMREFEFGSGMRRALAMSRGLRARAAGLGAELTDAQI--- 311  
DB 256 -LDMQLADSDAGEEDDAVLVARMBREYEVGSXLRN-SWSRG--AKRALIDYIDADVRRE 311  
QY 312 VEOE-----ESAPVWALIIPARSMNT 335  
DB 312 MEELIYKLAGLEAPERVSTVVAVALVAPDDMKLIQS 348

RESULT 6  
US-10-007-527A-24  
Sequence 24, Application US/10007527A  
Publication No. US20030044807A1  
GENERAL INFORMATION:  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Bramucci, Michael G.  
APPLICANT: Cheng, Qiong  
APPLICANT: Kostichka, Kristy N.  
TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors  
FILE REFERENCE: C11709 US NA  
CURRENT APPLICATION NUMBER: US/10/007,527A  
CURRENT FILING DATE: 2001-12-05  
PRIOR APPLICATION NUMBER: 60/254,868  
PRIOR FILING DATE: 2000-12-12  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 24  
LENGTH: 451  
TYPE: PRF  
ORGANISM: Streptomyces nigritifaciens  
US-10-007-527A-24

Query Match 16.1%; Score 322; DB 15; Length 451;  
Best Local Similarity 30.8%; Pred. No. 6e-24;  
Matches 101; Conservative 41; Mismatches 110; Indels 76; Gaps 15;

QY 49 PISGV-----NGVTIYNGPKSGFGGLRSCGKMWICPCGAGVGAHRADEISQVVAHOL 102  
DB 3 PASGVYVAQTAAGTSVVL-----GLMRCGRIMLCPVCAATIBHKRAEITAAVWEI 54  
QY 103 GTGSAVA-WVTMTMRHTAGORLHDLMTGLSAAMKAAATNGRR-----WRTE----- 145  
DB 55 KRGGTAYLVTFARIGHHTDLADLMDALQSTKTKDADPRPQAYORLITGGTWAGRRAYD 114  
QY 146 -----REMYCDGYVAVEITHG-KNGMHVHVALIMFSGDV-----SENILIESF 189  
DB 115 GHRADRBGIRDRIGYVGMIRATVTVQINGMHPIHIALVVGRTBESBAKQVGVF 174  
QY 190 --SDAMPRTSTSLVSLGPAAPLR-----NSGGLDVKRIIGGEAD-QVLAAYITKTA 237  
DB 175 BPSEALDDEWQOMPAV-WTAAIRKVNQPTDPDRHGVDFKLETERDANDLAAYAKTQ 233  
QY 238 SG--VGMREVSGDCKSGRGNRAPMEI--AYDAVG--DPA-----LELMREE 281  
DB 234 DKAAPALREARADLTKANGVNAPELLGRIGDLTGGMTEDDAAGVSGILEMMLARWHEE 293  
QY 282 FSGMGRALAMSRGLRARAAGLGAELTDA 309  
DB 294 RATKGRRALEWTRYLROMLGDGDPTEA 321



```

RESULT 7
US-10-007-452-24
; Sequence 24, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/1254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Streptomyces nigritfaciens
US-10-007-452-24

```

```

Query Match          16.1%; Score 322; DB 15; Length 451;
Best Local Similarity 30.8%; Pred. No. 6e-24;
Matches 101; Conservative 41; Mismatches 110; Indels 76; Gaps 15;

QY 49 PISGV-----NGVTVNGPKSGGFGLSGCKGMTCPCAGKVGARDEISQVYAHQ 102
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 3 PASGVVAQTAAAGTSVYL-----GLMRCGRIMVCPVCAATIRHRAEETAAVVEWI 54

QY 103 GGGGSA-VYTMTRHRAAGRIHDLMTGLSAMKATNGRR-----WRTE----- 145
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 55 KRGTVLYVTFTARHGHHTDLADLMDLQGTCKTAPAPRPAGYQRLTGGTYAGRAAD 114

QY 146 -----REMYGCDGYRAVEITHGK-NGMWHVHVALMFSGDV-----SENILSEF 189
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 115 GHRADREGIRDIQVGMIRAEVYVQGLNGHPIHIALVVGSTBGRSAKOIVGF 174

QY 190 --SDAMFDRWTSKLVLGFAPIR-----NSGLDVKTIQGEAD-QVLAAYITKIA 237
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 175 EPSEALDEWQGWRAV-WTAAARKNVQPTPDGRHGVFKLETERDANDLALEYIAKQ 233

QY 238 SG--VGMEVSGSGKSGRHNRAPIWEI--AVDAVG--DPA-----LELWREFE 281
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 234 DGAPALELAPADLKTANGNVAPFELIGRIQDLTGATWEDDAAGVGLSEWNLARWHEIE 293

QY 282 FGSGRRRAIWSRGLRPARAGIAGELTDA 309
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 294 RATKGRRAIEMTYILQMLGLDGDPTTA 321

```

```

RESULT 8
US-09-835-381-8
; Sequence 8, Application US/09835381
; Patent No. US20020045223A1
; GENERAL INFORMATION:
; APPLICANT: SUGA, Mikiko
; APPLICANT: ASAKURA, Yoko
; APPLICANT: MORI, Yukiko
; APPLICANT: ITO, Hisao
; APPLICANT: KURASHI, Osamu
; TITLE OF INVENTION: ARGININE REPRESSOR DEFICIENT STRAIN OF CORYNEFORM BACTERIUM AND M
; FILE REFERENCE: 206018USO
; CURRENT APPLICATION NUMBER: US/09/835,381
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: JP2000-129167
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 8
; LENGTH: 427

```

```

; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
US-09-835-381-8
Query Match          15.9%; Score 318; DB 9; Length 427;
Best Local Similarity 28.8%; Pred. No. 1.4e-23;
Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

QY 34 LQGITSETFNACR-----PISGVNGTVNGPKSGGFGLSGCKGMTCPCAGKV 86
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 27 MYKITNSKALAGCHRRMRDEAVNSMS-----NG--ASQPEGIQNHSRSGSLAELEV 79

QY 87 GAHRADEISQVYAHQIGTSVAVYTM-TWRTAGQRLHDLMTGLSAMKATNGRRMTE 145
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 80 MGERIELAIAIKHHLAAGALMFPVGVRRHRSQSFQVBAIGITAYVSMVKTSQMKKE 139

QY 146 REMYGCDGYRAVEITHG-KNGMWHVHVALMFSDVESENILSEFSDAMPDRWTSKLYSL 204
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 140 RARYGVHTYSDEVETDWMANGWHLRHMLFLRPLSDDEKAFEDSMFSRSGAVVKA 199

QY 205 GPAPLRNSG-GLD-VRKIGEADQVLAAYITKIASGVMEVSGDGKSGRHNRAPIWEI 262
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 200 GMDAPLRHGVKLDQVSTWGGDAK-MATYIAK--GMSQELTGSATITASKSYTPFGM 255

QY 263 AYDAVGDPQALE-----LWREFEFGSGRRRAIWSRGLRPARAGIAGELTDAQI--- 311
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 256 -LDMADQSDAGEDMDAVLVARWREYVSGSKILRS-SWSRG--AKRALGIDYIDADVRE 311

QY 312 VEOE-----ESAPYVVAIIIPASMMIET 335
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 312 MEELKYKLAGLEAPRVESTRVAVALVRPDDMKLIQS 348

```

```

RESULT 9
US-10-196-232-7
; Sequence 7, Application US/10196232
; Publication No. US20030113899A1
; GENERAL INFORMATION:
; APPLICANT: YAMAGUCHI, MIKIKO
; APPLICANT: ITO, HISAO
; APPLICANT: GUNII, YOSHITA
; APPLICANT: YASUEDA, HISASHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-ARGININE
; FILE REFERENCE: 225391USO
; CURRENT APPLICATION NUMBER: US/10/196,232
; CURRENT FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: JP 2001-224586
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 7
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum
US-10-196-232-7
Query Match          15.9%; Score 318; DB 15; Length 427;
Best Local Similarity 28.8%; Pred. No. 1.4e-23;
Matches 97; Conservative 58; Mismatches 132; Indels 50; Gaps 15;

QY 34 LQGITSETFNACR-----PISGVNGTVNGPKSGGFGLSGCKGMTCPCAGKV 86
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 27 MYKITNSKALAGCHRRMRDEAVNSMS-----NG--ASQPEGIQNHSRSGSLAELEV 79

QY 87 GAHRADEISQVYAHQIGTSVAVYTM-TWRTAGQRLHDLMTGLSAMKATNGRRMTE 145
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 80 MGERIELAIAIKHHLAAGALMFPVGVRRHRSQSFQVBAIGITAYVSMVKTSQMKKE 139

QY 146 REMYGCDGYRAVEITHG-KNGMWHVHVALMFSGDVSENILSEFSDAMPDRWTSKLYSL 204
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 140 RARYGVHTYSDEVETDWMANGWHLRHMLFLRPLSDDEKAFEDSMFSRSGAVVKA 199

QY 205 GPAPLRNSG-GLD-VRKIGEADQVLAAYITKIASGVMEVSGDGKSGRHNRAPIWEI 262
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

Db      200 GMDAPLREHGVLTQVSTGSDAK--MATYLAK---GMSGLTSATKASGSTTTPGM 255
      263 AVDAVGGDPALE-----LMREFEGSGRRALAMSRGLPARAGAEITDAQI--- 311
      256 -IDMLADQSDAGEDMDVALVARREVEVSGKNLRS--SMERG--AKRALIDYIDADVRE 311
Qy      312 VEOE-----ESAPVWVAIIPARSMRT 335
      312 MEEELYKLAGLEAPERVESTRAVALVKPDWKLIQS 348

```

RESULT 10  
US-10-007-527a-22

```

; Sequence 22, Application US/10007527A
; Publication No. US20030044807A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.
; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,527A
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Streptomyces lividans
US-10-007-527a-22

```

Query Match 15.8%; Score 316; DB 15; Length 456;  
Best Local Similarity 30.5%; Pred. No. 2.5e-23;  
Matches 100; Conservative 40; Mismatches 112; Indels 76; Gaps 15;

```

Qy      49 PISGV-----NGVTIVNGPKSGSGFGLRSCGKGMICPCGAKVGAHRADEISQVAAHQL 102
      3 PASGVIYAQTAAAGTSVVL-----GLMRCGRIMLCPVCAATIRHRAEITAAVEMVI 54
Db      103 GTGSVA-WVTMTMHTAGORLHDLMTGLSAAMKAATNGRR-----WTE----- 145
      55 KKGTAIYVTFARHGHTRDLADLMDALQSTRKTPSPRRPAGAYQRLITGTMAGRRAKD 114
      146 -----REMYGCDGVRAVEITTHGK--NGMHVYHALMFSGDV-----SENTLESEF 189
      115 GHRADREGIRIDRIGVGMIRATEVTVQINGMHPHIAVLVGGRTGERSAKQIVATF 174
      190 --SDAMFDRWTSKLVSLSGFAPLR-----NSGDLVRKIGGAD-QVLAAYLTXTA 237
      175 EPTGALLDEMGHMSV-WTAAIRKVNPAFTPDHGRGVDFKLETERDANDLAEYIAKQ 233
      238 SG--VGMVEVSGDGSGRGNRAPEI--AVDAVGG--DPOA-----LELMREFE 281
      234 DGKAPALBELARADLKTATGCVNAPFELIGRIDLTGGMTEDDAAGVSGLEMLSRMHEYE 293
Qy      282 FSGMGRRAIAMSRLPARAGAEITDA 309
      294 RATRGRAIEMTRYLRQMLGLDGDTEA 321
Db

```

RESULT 11  
US-10-007-452-22

```

; Sequence 22, Application US/10007452
; Publication No. US20030093701A1
; GENERAL INFORMATION:
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Bramucci, Michael G.
; APPLICANT: Cheng, Qiong
; APPLICANT: Kostichka, Kristy N.

```

```

; TITLE OF INVENTION: Rhodococcus Cloning and Expression Vectors
; FILE REFERENCE: CL1709 US NA
; CURRENT APPLICATION NUMBER: US/10/007,452
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: 60/254,868
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 22
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Streptomyces lividans
US-10-007-452-22

```

Query Match 15.8%; Score 316; DB 15; Length 456;  
Best Local Similarity 30.5%; Pred. No. 2.5e-23;  
Matches 100; Conservative 40; Mismatches 112; Indels 76; Gaps 15;

```

Qy      49 PISGV-----NGVTIVNGPKSGSGFGLRSCGKGMICPCGAKVGAHRADEISQVAAHQL 102
      3 PASGVIYAQTAAAGTSVVL-----GLMRCGRIMLCPVCAATIRHRAEITAAVEMVI 54
Db      103 GTGSVA-WVTMTMHTAGORLHDLMTGLSAAMKAATNGRR-----WTE----- 145
      55 KKGTAIYVTFARHGHTRDLADLMDALQSTRKTPSPRRPAGAYQRLITGTMAGRRAKD 114
      146 -----REMYGCDGVRAVEITTHGK--NGMHVYHALMFSGDV-----SENTLESEF 189
      115 GHRADREGIRIDRIGVGMIRATEVTVQINGMHPHIAVLVGGRTGERSAKQIVATF 174
      190 --SDAMFDRWTSKLVSLSGFAPLR-----NSGDLVRKIGGAD-QVLAAYLTXTA 237
      175 EPTGALLDEMGHMSV-WTAAIRKVNPAFTPDHGRGVDFKLETERDANDLAEYIAKQ 233
      238 SG--VGMVEVSGDGSGRGNRAPEI--AVDAVGG--DPOA-----LELMREFE 281
      234 DGKAPALBELARADLKTATGCVNAPFELIGRIDLTGGMTEDDAAGVSGLEMLSRMHEYE 293
Qy      282 FSGMGRRAIAMSRLPARAGAEITDA 309
      294 RATRGRAIEMTRYLRQMLGLDGDTEA 321
Db

```

RESULT 12  
US-09-826-191-9

```

; Sequence 9, Application US/09826191
; Patent No. US2002000665A1
; GENERAL INFORMATION:
; APPLICANT: Stoddard, Steven F.
; TITLE OF INVENTION: KETOGLUCONITUM ENDOGENOUS PLASMIDS
; FILE REFERENCE: 1533.1260001
; CURRENT APPLICATION NUMBER: US/09/826,191
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: US 60/194,627
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Ketoglucanigenium
US-09-826-191-9

```

Query Match 14.1%; Score 282; DB 9; Length 466;  
Best Local Similarity 30.0%; Pred. No. 7.2e-20;  
Matches 87; Conservative 40; Mismatches 119; Indels 44; Gaps 15;

```

Qy      60 NGPKSGFGLRSCGKGMICPCGAKVGAHRADEISQVAAHQLGTGSV-AMVTMTMHTA 118
      109 DGGRRAFFSGLAQCRNVGCAVCSARIAQIRSEMMHLLAWARDNGFVPLITLTLQHKR 168
      119 GQRHDLMTGLSAAMKAATNGRRMTEREMYGCD---GVYRAVEITTHG-KNGMHVYHVA 173

```

Db 169 GDSLFDDLQNMKKAQRLRORREWR-----DLPEVGSVTSTETHSYANGMHPHFE 220  
Qy 174 L-LMFSGDVSE--NILESFDAMFDRWTSKLVISGFAPLPLNSGGLDVKXIGEGADQVLA 230  
Db 221 IVLRAAGDESEALHLMQRLGDA---WRACLKGYGMWG---NDAAPDRGAANGD----- 269  
Qy 231 AYLTKIASGVMEVSGDGSGRHNRAPEIAVDVAGDPOALEMREFFGSGWG--RR 288  
Db 270 -YVAK--WGAEEELTSSSKSGKRGTRPROLL--QAQDD---GLMIEFNATSGKRR 320  
Qy 289 AIAWSRGLRARAAGLAEITD---AQIVEQESAPVWVAIIPASMMIR 334  
Db 321 QLVMSQGLKEEGCL-VELDDDEAMAEVDAEAGSGPEIYAEMDNBGMKQVR 369

RESULT 13  
US-10-263-666-9  
; Sequence 9, Application US/10263666  
; Publication No. US20030073224A1  
; GENERAL INFORMATION:  
; APPLICANT: D'Elia, John  
; APPLICANT: Stoddard, Steven F.  
; TITLE OF INVENTION: KETOGLONIGENIUM ENDOGENOUS PLASMIDS  
; FILE REFERENCE: 1533.1260004  
; CURRENT APPLICATION NUMBER: US/10/263,666  
; PRIOR FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: US 09/826,191  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: US 60/194,627  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Ketogulonigenium  
US-10-263-666-9

Query Match 14.1%; Score 282; DB 15; Length 466;  
Best Local Similarity 30.0%; Pred. No. 7.2e-20;  
Matches 87; Conservative 40; Mismatches 119; Indels 44; Gaps 15;  
Qy 60 NPKSGGFGGLRSCGKGMICPCGAGKVGARADEISQVVAHQLTGTSV-AMVTMTMRHTA 118  
Db 109 DGGRRAFFSGLAQCRRNWGCVCASARIADIRSENMHLLAMARDNGFVPLITLTAQHKA 168  
Qy 119 GQRLHDLMTGLSAAMKAAATNGRRMTEREMYGCD---GYRAVEITHG-KXGMHVHYHA 173  
Db 169 GDSLFDDLQNMKKAQRLRORREWR-----DLPEVGSVTSTETHSYANGMHPHFE 220  
Qy 174 L-LMFSGDVSE--NILESFDAMFDRWTSKLVISGFAPLPLNSGGLDVKXIGEGADQVLA 230  
Db 221 IVLRAAGDESEALHLMQRLGDA---WRACLKGYGMWG---NDAAPDRGAANGD----- 269  
Qy 231 AYLTKIASGVMEVSGDGSGRHNRAPEIAVDVAGDPOALEMREFFGSGWG--RR 288  
Db 270 -YVAK--WGAEEELTSSSKSGKRGTRPROLL--QAQDD---GLMIEFNATSGKRR 320  
Qy 289 AIAWSRGLRARAAGLAEITD---AQIVEQESAPVWVAIIPASMMIR 334  
Db 321 QLVMSQGLKEEGCL-VELDDDEAMAEVDAEAGSGPEIYAEMDNBGMKQVR 369

; CURRENT APPLICATION NUMBER: US/10/261,481  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: US 09/826,191  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: US 60/194,627  
; PRIOR FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Ketogulonigenium  
US-10-261-481-9

Query Match 14.1%; Score 282; DB 15; Length 466;  
Best Local Similarity 30.0%; Pred. No. 7.2e-20;  
Matches 87; Conservative 40; Mismatches 119; Indels 44; Gaps 15;  
Qy 60 NPKSGGFGGLRSCGKGMICPCGAGKVGARADEISQVVAHQLTGTSV-AMVTMTMRHTA 118  
Db 109 DGGRRAFFSGLAQCRRNWGCVCASARIADIRSENMHLLAMARDNGFVPLITLTAQHKA 168  
Qy 119 GQRLHDLMTGLSAAMKAAATNGRRMTEREMYGCD---GYRAVEITHG-KXGMHVHYHA 173  
Db 169 GDSLFDDLQNMKKAQRLRORREWR-----DLPEVGSVTSTETHSYANGMHPHFE 220  
Qy 174 L-LMFSGDVSE--NILESFDAMFDRWTSKLVISGFAPLPLNSGGLDVKXIGEGADQVLA 230  
Db 221 IVLRAAGDESEALHLMQRLGDA---WRACLKGYGMWG---NDAAPDRGAANGD----- 269  
Qy 231 AYLTKIASGVMEVSGDGSGRHNRAPEIAVDVAGDPOALEMREFFGSGWG--RR 288  
Db 270 -YVAK--WGAEEELTSSSKSGKRGTRPROLL--QAQDD---GLMIEFNATSGKRR 320  
Qy 289 AIAWSRGLRARAAGLAEITD---AQIVEQESAPVWVAIIPASMMIR 334  
Db 321 QLVMSQGLKEEGCL-VELDDDEAMAEVDAEAGSGPEIYAEMDNBGMKQVR 369

RESULT 15  
US-10-261-942-9  
; Sequence 9, Application US/10261942  
; Publication No. US20030087440A1  
; GENERAL INFORMATION:  
; APPLICANT: D'Elia, John  
; APPLICANT: Stoddard, Steven F.  
; TITLE OF INVENTION: KETOGLONIGENIUM ENDOGENOUS PLASMIDS  
; FILE REFERENCE: 1533.1260002  
; CURRENT APPLICATION NUMBER: US/10/261,942  
; PRIOR FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: US 09/826,191  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: US 60/194,627  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Ketogulonigenium  
US-10-261-942-9



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 22:29:38 ; Search time 34 Seconds  
(without alignments)  
1071.998 Million cell updates/sec

Title: US-10-007-527A-2

Perfect score: 2005

Sequence: 1 MTSVSAEHLGKDRPPVIVS.....HLHYRLPADVPRPIISVRK 379

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321.5	16.0	256	S34850	hypothetical prote
2	316	15.8	456	A31844	rep protein - Stre
3	230.5	11.5	437	S04020	hypothetical prote
4	169	8.4	240	T40335	hypothetical prote
5	127	6.3	314	B35390	replication protei
6	123.5	6.2	315	S25721	replication protei
7	122.5	6.1	315	T44800	replication protei
8	118.5	5.9	315	JN0856	rep protein - Str
9	115	5.7	336	A40651	replication protei
10	114	5.7	314	T46831	plasmid replicatio
11	111.5	5.6	333	A49783	replication initia
12	105	5.2	381	A83561	probable type II s
13	104.5	5.2	340	T140561	replication initia
14	104	5.2	420	AG2059	heme biosynthesis
15	102.5	5.1	420	AC3284	D-amino-acid dehyd
16	101.5	5.1	339	S01098	rep protein - Bact
17	100.5	5.0	614	G75393	hypothetical impo
18	99.5	5.0	614	D87410	fold protein [impo
19	98.5	4.9	532	H83493	hypothetical prote
20	98.5	4.9	353	E83641	probable hemagglut
21	98	4.9	442	D87594	genetic exchange p
22	95.5	4.8	606	D83484	probable acyl-CoA
23	94.5	4.7	1198	T28678	polyketide synthas
24	94.5	4.7	1763	T17465	ribose polymerase
25	94	4.7	317	A43621	probable replicati
26	94	4.7	382	A82861	L-lactate dehydrog
27	94	4.7	382	D76338	chaperonin groEL
28	94	4.7	544	BVTCGL	hypothetical prote
29	94	4.7	1245	H87340	hypothetical prote

30	93	4.6	317	2	S25294	replication protei
31	93	4.6	748	2	T37097	probable secreted
32	93	4.6	796	2	S57844	lethal (3) malignant
33	92.5	4.6	354	2	B75355	hypothetical prote
34	92.5	4.6	600	2	E72027	phosphoenolpyruvat
35	92.5	4.6	600	2	A86597	phosphoenolpyruvat
36	92.5	4.6	1026	2	G87346	hypothetical prote
37	92	4.6	799	2	T48889	serine/threonine p
38	91.5	4.6	302	2	H96792	unknown protein F1
39	91.5	4.6	332	2	S51694	replication protei
40	91.5	4.6	350	2	B82281	ferric vibriobacti
41	91	4.5	296	1	CPNB7	peroxidase (EC 1.1
42	91	4.5	956	2	B83200	probable phosphotr
43	90.5	4.5	320	2	D96750	unknown protein F2
44	90.5	4.5	2399	2	AH3009	non-ribosomal pept
45	90.5	4.5	2399	2	F98274	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S34850  
hypothetical protein - Streptomyces lavendulae plasmid pSLG33

C/Species: Streptomyces lavendulae

C/Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 07-May-1999

C/Accession: S34850

R/Felsberg, J.; Petricek, M.; Tichy, P.

Nucleic Acids Res. 21, 3582, 1993

A/Title: Nucleotide sequence of the mini-plasmid pSLG33 from Streptomyces lavendulae

A/Reference number: S34850; MUID:93346001; PMID:8346038

A/Accession: S34850

A/Status: preliminary; translation not shown

A/Molecule type: DNA

A/Residues: 1-256 <FEEL>

A/Cross-references: EMBL:X69872

A/Genetics:

A/Genome: plasmid pSLG33

Query Match 16.0%; Score 321.5; DB 2; Length 256;

Best Local Similarity 33.2%; Pred. No. 5.3e-19;

Matches 85; Conservative 45; Mismatches 105; Indels 21; Gaps 11;

QY	109	WVTMTTRHTAQQRLDMLWTGISAANKAATNGRRM-----TERENYGGDYRAVEI	160
DB	1	WVTTLARHKKHHELEPLDAVANGWRKLLSGRAWAGDPKRGVLAGRDRLGYRGNIRSLV	60
QY	161	THG-KNGHNVVHALMFGSDVSENIIESPSDAMPDR-WTSKTVSLGFAAPLRNSGGDV	218
DB	61	TYGTNGMHPHLVLLINBETTEL--AYAMHRWDKTRAWLKKAGPEPSKEH--GITW	116
QY	219	RKIGEAQOVLAAVYTKIASG-VGMEVSGSGKSGRGNRAPWE-IAVDVAGDQAL	276
DB	117	SKV--TTPEBAGEYIAKQBGKVGNEIRAGDKKARLGTLPFLETFRRQGDVAVPI	174
QY	277	WREFFSGMGRRAIWSRGLRPAR-AGLGAELTDAQIVQESAPVVAIIPARSMWMT	335
DB	175	WQEVKGTFRRAITWSRGLRAELLDGDEBELDAEELIGGFTW-ALLPESRAIR	233
QY	336	CAPYVFGELIGLVENG	351
DB	234	-VPGIQSRILDTAENG	248

##### RESULT 2

rep protein - Streptomyces lividans plasmid pJ101

C/Species: Streptomyces lividans

C/Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 23-Mar-1993

C/Accession: A31844

R/Kendall, K.J.; Cohen, S.N.

J. Bacteriol. 170, 4634-4651, 1988

A/Title: Complete nucleotide sequence of the Streptomyces lividans plasmid pJ101 and

A:Reference number: A91888; MUID:89008081; PMID:3170481  
A:Accession: A31844  
A:Molecule type: DNA  
A:Residues: 1-456 <KEN>  
C:Genetics:  
A:Genome: plasmid

Query Match 15.8%; Score 316; DB 2; Length 456;  
Best Local Similarity 30.5%; Pred. No. 2.9e-18;  
Matches 100; Conservative 40; Mismatches 112; Indels 76; Gaps 15;

QY 49 PISGV-----NGVTIVNGPKSGFGGLRSCGKMTCPCCAGKAGHRADEISOVVAHQ 102  
DB 3 PASGIVAGTAAGTAVVL-----GLMRCGRIMLCVCAATIRHKRAEETITAAVWMI 54  
QY 103 GTGSVA-MVTMTMTTAQORLHDLMTGSAAMKATNRR-----WRTE----- 145  
DB 55 KRGGTAYLVTFRTARIGHTRLADLMDALQGTTRKTPDSPRRPAGVQRLITGGTWAGRRAKD 114  
QY 146 -----REMYGCDGVRAVEITTHGK-NGMHVHALIMFSGDV-----SENILRSF 189  
DB 115 GHRADREGIRDRIGVGMIRATEVTVGQINGMHPIHALIVVGRTEGERSAKQIVATF 174  
QY 190 --SDMPFRWTSKVLSPAPAPLR-----NSGGLDVKRKIGSEAD-QVLAAYLTIA 237  
DB 175 EPTGAALDEWQGHMRSV-WTAALRKVNPAFTPDHGHGDFKRLETERDANDIAEYIAQTQ 233  
QY 238 SG-VGMEVSGDGSGRHGNAPMEI--AYDAVGS--DPQA-----LELMREFE 281  
DB 234 DKKAPALBLARADLKTATGCVNAPPELLGRIGDLTGWTEDDAAGVSLNNLSRMHEYE 293  
QY 282 FSGMGRRAIAMSRLGARAPAGLGAELTDA 309  
DB 294 RATRGRRALIEWTRYLRQMLGLDGGDTTEA 321

## RESULT 3

S04020  
hypothetical protein - Streptomyces sp. plasmid PSB24.2

C:Species: Streptomyces sp.  
C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 28-Oct-1994  
C:Accession: S04020  
R:Boletini, A.P.; Sorokin, A.V.; Aleksandrov, N.N.; Danilenko, V.N.; Kozlov, Y.I.  
Dokl. Biochem. 283, 260-263, 1985  
A:Title: Nucleotide sequence of DNA of the actinomycete plasmid PSB24.2.  
A:Reference number: S04020  
A:Accession: S04020  
A:Molecule type: DNA  
A:Residues: 1-437 <BOU>  
A:Cross-references: EMBL:X03756

A:Note: the authors translated the codon TGG for residue 59 as Arg, CGG for residue 187  
A:Note: the source is designated as Streptomyces cyanogenus  
A:Note: this paper is a translation of the Russian paper published in Dokl. Akad. Nauk S  
C:Genetics:  
A:Genome: plasmid

Query Match 11.5%; Score 230.5; DB 2; Length 437;  
Best Local Similarity 24.6%; Pred. No. 3.1e-11;  
Matches 102; Conservative 43; Mismatches 140; Indels 129; Gaps 19;

QY 55 GVTIVNGPKSGFG-GLRSCGKMTCPCCAGKAGHRADEISOVVAHQLTGSVA-MVTM 112  
DB 6 GVLFARTAGAAVAAGLILKCGRIWLCFVCSGGDPAPRSEITEAVVSWLQGGGMAYLVTF 65  
QY 113 TMRHTAGORLHDLMTL-----SAAWKATNGRRWRTEREY 149  
DB 66 TARHTAADRLSDMLALQGTADATGTRRRPAGVQRLITGGAM--AGDKRRKSNQEGIR 123  
QY 150 GCDGY---VRAVEITTHGK-NGMHVHALIMFSGDV-----SENILRSF---DAF--- 194  
DB 124 GATIGYGMTIRATEVTYGBGAGMHPHALIVVGRTEGERSAKQIVATF 183  
QY 195 DRWTSKVLVSLGPAAPLRNSGL---DVARKIGSEADQVLAAYLTIASGVG----- 241

DB 184 DRMRRS-----GPATLASTGFPRTGARSPPGATGATATGSSSSVRSGRQPPGVHR 239  
QY 242 -----MEVSGDGKSGRHN-----RAPWEIADVAGDPO-----ALELW 277  
DB 240 QDAGROEPFGAGARRRQGRPPQGHVLTFRTPSRIG-DLMGVBEELAGSLAWGLDRW 298  
QY 278 REEFEGSG-----RRALMSRGL-----RAPAGL 302  
DB 299 AEYETAVSAGAHVDPPLPAPGPDRRRHGRHRRGPPVPRRRRRFRDGVQIMDRAMWL 358

QY 303 GAELTDAQVIEQESAEVWVAII-----PARSMMTRTCAPY-----VFGELIG 346  
DB 359 VGRSLDAVEAEVAGREISMDALGELVQASQSAFLRVLTPOEVELYELLG 412

## RESULT 4

140335  
hypothetical protein 1 - Corynebacterium glutamicum

C:Species: Corynebacterium glutamicum  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 22-Oct-1999  
C:Accession: 140335  
R:Yamauchi, R.; Terabe, M.; Miwa, K.; Tsuchiya, M.; Takagi, H.; Morigata, Y.; Nakamori  
Agric. Biol. Chem. 50, 2771-2778, 1986  
A:Title: Determination of the complete nucleotide sequence of Brevibacterium lactoferm  
A:Reference number: 140334  
A:Accession: 140335  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-240 <RS>  
A:Cross-references: GB:D00038; NID:g216381; PIDN:BAA0026.1; PID:g216383

Query Match 8.4%; Score 169; DB 2; Length 240;  
Best Local Similarity 32.0%; Pred. No. 1.8e-06;  
Matches 54; Conservative 28; Mismatches 53; Indels 34; Gaps 10;

QY 193 MFRWTSKVLVSLGPAAPLRNSG-GLD-VRKIGSEADQVLAAYLTIASGVMEVSGSGK 250  
DB 1 MFSRMSAGVAVAGMDAPLREHGVKLDQVSTGGDAAK-MATYIAK---GMSQELTGSATK 56  
QY 251 SGRHGNRAPMEIADVAGDPOALE-----LMREFFSGMSGRALAMSRLGARAGL 302  
DB 57 TASKSSTYPPQM-IDMLADQSDAGBMDAVYAKRREIVSGSKURS-SMSRG-AKDAL 112  
QY 303 GAELTDAQI---VEQE-----ESAPVVAIIPARSMMTRT 335  
DB 113 GIDYIDAVRREMEBELYKLAGLAPKRVSTRAVALVKRDDMTLIGS 161

## RESULT 5

B35390  
replication protein - Lactobacillus hilgardii

C:Species: Lactobacillus hilgardii  
C:Date: 31-Aug-1990 #sequence\_revision 31-Aug-1990 #text\_change 24-Sep-1999  
C:Accession: B35390  
R:Jooson, K.; Soetaert, P.; Michiels, F.; Joos, H.; Mahillon, J.  
J. Bacteriol. 172, 3089-3099, 1990  
A:Title: Lactobacillus hilgardii plasmid pLAB1000 consists of two functional cassette  
A:Reference number: A35390; MUID:90264300; PMID:2188951  
A:Accession: B35390  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-314 <JOS>  
A:Cross-references: GB:M55222; NID:g149352; PIDN:AAA98163.1; PID:g149354; GB:M37906

C:Superfamily: replication protein

Query Match 6.3%; Score 127; DB 2; Length 314;  
Best Local Similarity 22.0%; Pred. No. 0.0071;  
Matches 55; Conservative 50; Mismatches 107; Indels 38; Gaps 10;

QY 75 KGMICPPCAGKAGHRADEISOV--AH-QLGTGSVAWMTMTTAAQORLHDLMTGSA 131  
DB 69 KSRICPLCNWRSGSQSNQMLQVLDPAIKQRTKGFPLUTITAEASGENLKQVRYKMR 128

```

RESULT 7
T44800
replication protein repA [imported] - Streptococcus thermophilus plasmid pER35
C:/Species: Streptococcus thermophilus
C:/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C:/Accession: T44800
R:/SOLow, B.T.; Somkuti, G.A.
submitted to the EMBL Data Library, August 1999
A:/Description: Comparison of low molecular weight heat stress proteins encoded on plasmid
A:/Reference number: Z22847
A:/Accession: T44800
A:/Status: preliminary; translated from GB/EMBL/DBJ
A:/Molecule type: DNA
A:/Residues: 1-315 <SOL>
A:/Cross-references: EMBL:AF177167, PIDN:AA04355.1
A:/Experimental source: strain ST135
A:/Genetics:

```

[illegible]

RESULT 9  
A40651  
replication protein A - *Synechocystis* sp. plasmid pCA2.4  
CISpecies: *Synechocystis* sp.  
C.Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 24-Sep-1999  
C.Accession: A40651  
R.Yang, X.; McFadden, B.A.  
J. Bacteriol. 175, 3981-3991, 1993  
Article: A small plasmid, pCA2.4, from the cyanobacterium *Synechocystis* sp. strain PC



N;Contains: DNA ligase (EC 6.5.1.-); DNA lyase (EC 4.2.99.-)

```

QY      7  WICPCAGVGNH--RADISQVVAHOGTGSVAMVTMTHTAGORLHDLMTGLSAAK 134
DQ      127  WV---AGWDGHLIRNSLDVAVHPLEOGLEVP-----QELRFLWIGBERPAK 174
QY      135  AATNGR-----RWRTREMYGCGDYVAWEILTHKNGWH-----VHYALL 175
DQ      175  SAALREAFALDEALRWSPGPWGL--ALKAPATPSAGNGVTCKMPLFCALALLVWYAGAN 232
QY      176  MEGDVSEVNIIESFSPAMFDRWTSKLYNSLGPAA-----PIRNGGGLDKIGGBADYL 229
DQ      233  LVYGOALF-----QGSLQRQSSQRYOAAFEPELVPLDPIKQREBRDAYLNGKADGA 286
QY      230  AAYLTIASGVGMEVSGBDKSGR---HGNRAPMEIADVAGDPQALREPREPFGSWG 286
DQ      287  APGLAALLHGAGEAMFLAGRLQRLDYHAGELDELPLGFLPGGAAA---W-QGELKSG 342
QY      287  RRALWRSGLRARA 300
DQ      343  IQADADCKGMQVRA 356

```

RESULT 13  
 140561  
 replication initiation protein rep60 [validated] - Bacillus subtilis plasmids  
 N:Contains: DNA ligase (EC 6.5.1.-); DNA lyase (EC 4.2.99.-)  
 C:Species: Bacillus subtilis  
 C:Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text\_change 20-Apr-2000  
 C:Accession: I40561, A32059, A43997, I40549  
 R:Weijer, W.J.; Venema, G.; Bion, S.  
 Nucleic Acids Res. 23, 612-619, 1995  
 A:Title: Characterization of single strand origins of cryptic rolling-circle plasmids from  
 A:Reference number: I40549; MUID:95206941; PMID:7899081  
 A:Accession: I40561  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-340 <MEI>  
 A:Cross-references: EMBL:U3380; NID:G1049123; PID:AA044418; PID:G1049124  
 A:Experimental source: plasmid pTA1060  
 R:Devine, K.M.; Hogan, S.T.; Higgins, D.G.; McConnell, D.J.  
 J. Bacteriol. 171, 1166-1177, 1989  
 A:Title: Replication and segregation stability of Bacillus plasmid pBA41.  
 A:Reference number: A32059; MUID:89123116; PMID:2492507  
 A:Accession: A32059  
 A:Molecule type: DNA  
 A:Residues: 34-124, 'V', 126-138, 'E', 140-230, 'L', 232-237, 'E', 239-340 <DEV>  
 A:Cross-references: GB:M24251; NID:G142426  
 A:Experimental source: plasmid pBA41  
 A:Note: the authors' translation is shown for residues 313-333 and not the sequence shown  
 R:Darabi, A.; Porough, R.; Bhatnagar, G.; Watabe, M.; Goodarzi, G.; Gross, S.C.; Watabe,  
 Plasmid 22, 281-286, 1989  
 A:Title: Identification and nucleotide sequence of the minimal replicon of the low-copy  
 A:Reference number: A43997; MUID:90222329; PMID:2561213  
 A:Accession: A43997  
 A:Molecule type: DNA  
 A:Residues: 34-124, 'V', 126-138, 'E', 140-230, 'L', 232-237, 'E', 239-301, 'R', <DAR>  
 A:Cross-references: GB:M24386; NID:G144127; PID:AAA23002.1; PID:G144129  
 A:Experimental source: plasmid pBS2  
 A:Note: the authors translated the codon ACG for residue 124 as Tyr, and GAG for residue  
 C:Genetics:  
 A:Gene: rep60  
 A:Genome: plasmid  
 A:Start codon: TTG  
 C:Superfamily: replication protein  
 C:Keywords: carbon-oxygen lyase; ligase; plasmid replication  
 C:Keywords: carbon-oxygen lyase; ligase; plasmid replication  
 F:252/Active site: Tyr (covalent DNA-binding) #status predicted  
 Query Match 5.2%; Score 104.5; DB 2; Length 340;  
 Best Local Similarity 24.2%; Pred. No. 0.56;  
 Matches 38; Conservative 35; Mismatches 57; Indels 27; Gaps 9;

[illegible]

RESULT 14  
AG2059  
heme biosynthesis protein [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AG2059  
R:Nakako, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriugu  
K.; Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:1595285; PMID:11759840  
A:Accession: AG2059  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-820 <KUR>  
A:Cross-references: GB:BA000015, PIDN:BAE73728.1, PID:G17131120, GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: al12029

```

Query Match      5.2%; Score 104; DB 2; Length 420;
Best Local Similarity 24.4%; Pred. No. 0.79;
Matches      85; Conservative 47; Mismatches 135; Indels 82; Gaps 23;

QY  Q0ITTSEFNACGPIS-GVNGVYIVNGPKSGGGGIRS-----CGKGIWPCGAC 84
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB  QELSTAEALNIVQGLAEVGIKEVTLIGGE-----AFLRPDMVIYAKITDAGMLCGMTTG 90
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY  KVGARADDEISQVVAHQIGGSVAMVTMTKRTNAG-QRLDMLTGLSAANKATNGRRMR 143
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB  GYG-----ISLENAQKRRKAGIAKSVS---TDGEMATDHLNLRGKSGWSGL----- 135
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY  TEREWYGGDGVRAVEITTHGKNGHVAHVALIMFSGDVSENILESFDAMFDRWTSXL-V 202
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB  L36--RIM---SYLKKVGLIFGCN---TQINL---SAPFSPSYIENHIDAGARMQIDLV 183
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY  203 SLGPAAP----LNSGGLDYRKIGERADVLAAVLTLYIASGVMEVSGDGKSGRRHNR 257
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB  184 PMGNADANADITLQPSSELDL-----YPMLARVQRANTGVRVAAGN-NIGYGG-- 232
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY  258 APMEIADVAVAGDPOALELMREFEFG--SMGRRAIAMSRG--LRAAVGLAELTD---A 309
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB  233 -PYE---RLRGGCKEMGFMKGCAGISTGLDELNDGRTIKCPSPITPAVYAGNIRERPLR 286
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY  310 QIVQESAPV-MVAIIP--ARSMMIRTCAPYVSEIILGLVEAGATW 354
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB  289 EIVEQSELEPNLYAGTPEGTAYLMDGCKTCK--FAE---LCRGGCTW 331
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 15
ACG284
D-amino-acid dehydrogenase (EC 1.4.99.1) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C:Accession: ACG284
R:DelVecchio, V.G.; Kapatkar, V.; Redkar, R.J.; Patra, G.; Wajer, C.; Los, T.; Ivanov, J.;
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AJ3252; PMID:1175688
A:Status: Preliminary

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 23:10:48 ; Search time 24 Seconds  
(without alignments)  
742.630 Million cell updates/sec

Title: US-10-007-527A-2  
Perfect score: 2005

Sequence: 1 MTSVSAEHLSEKDRPVLS.....HHRRLPADVPRPISVRK 379

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	316	15.8	456	1	REP_STRLI
2	227	11.3	437	1	NBGR_STRCY
3	127	6.3	314	1	REP_LACHT
4	102.5	5.1	309	1	REP_BACSP
5	101.5	5.1	339	1	REP_BACAM
6	99	4.9	317	1	REP_LACPL
7	95	4.7	697	1	PURL_PYRAE
8	94.5	4.7	431	1	ARGD_BIRFO
9	94	4.7	544	1	CH60_SYNF
10	93	4.6	796	1	MBN_DROME
11	92.5	4.6	600	1	PCCK_CHLPN
12	92	4.6	799	1	AFSK_STRCO
13	91	4.5	296	1	PERX_BRARA
14	90	4.5	385	1	RECF_MYCPA
15	90	4.5	571	1	FLBI_CAMEB
16	89.5	4.5	543	1	CH60_AQUAE
17	89.5	4.5	783	1	TRF2_THBAC
18	88.5	4.4	622	1	PPCC_HUMAN
19	88.5	4.4	664	1	DING_MYCTU
20	88	4.4	384	1	TGP_XYLPA
21	88	4.4	546	1	CH60_BURYI
22	87.5	4.4	692	1	AKA8_HUMAN
23	87	4.3	959	1	RBL_CALSH
24	87	4.3	954	1	FLFY_CAUCR
25	86	4.3	447	1	DHE4_CORGL
26	86	4.3	461	1	PSBC_CYAPA
27	86	4.3	645	1	DXS_RATME
28	85.5	4.3	659	1	MUTL_HALNI
29	85	4.2	385	1	AGAS_STRCO
30	85	4.2	546	1	CH60_BURCE
31	85	4.2	637	1	DXS_RHILC
32	85	4.2	1023	1	STRB_HUMAN
33	84.5	4.2	544	1	CH60_FRATU

34	84	4.2	322	1	PERX_NICSY	Q02200 nicotiana s
35	84	4.2	466	1	SCRB_SALTY	P37075 salmoneila
36	84	4.2	511	1	RHB1_YEAST	Q04225 saccharomyc
37	84	4.2	1334	1	JIP3_HUMAN	Q9up62 homo sapien
38	83.5	4.2	324	1	PS52_ARATH	Q9f1c0 arabidopsis
39	83.5	4.2	408	1	ASSY_HELMO	Q8gdu2 heliobacill
40	83.5	4.2	475	1	RBL_NOTSU	Q8gdu2 heliobacill
41	83.5	4.2	622	1	MAK_RAT	P20793 ratu
42	83	4.1	214	1	CPGF_SYNT3	P72652 synchocyst
43	83	4.1	522	1	GA6S_HUMAN	P34059 homo sapien
44	83	4.1	562	1	ILVB_BCOLI	P08142 escherichia
45	83	4.1	712	1	PAL2_PHAVU	P19142 phaseolus v

## ALIGNMENTS

RESULT 1	REP_STRLI	STANDARD;	PRT;	456 AA.
ID	REP_STRLI			
AC	P22406;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	01-AUG-1991 (Rel. 19, Last annotation update)			
DE	Rep protein.			
GN	REP.			
OS	Streptomyces lividans.			
OG	Plasmid p101.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomyces; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxId=1916;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89008081; PubMed=3170481;			
RA	Kendall K.J., Cohen S.N.;			
RT	"Complete nucleotide sequence of the Streptomyces lividans plasmid p101 and correlation of the sequence with genetic properties.";			
RL	J. Bacteriol. 170:4634-4651(1988).			
CC	-1- FUNCTION: ESSENTIAL FOR THE AUTONOMOUS REPLICATION OF THE PLASMID p101.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; M21778; AAAB8404.1; -			
DR	PIR; A31844; A31844.			
DR	InterPro; IPR000989; Rep.			
KW	Pfam; PF01446; Rep. 1;			
KW	plasmid; Plasmid partition.			
SQ	SEQUENCE 456 AA; 49617 MW; 0976FD9D9429C7DC CRC64;			
Query Match	15.8%; Score 316; DB 1; Length 456;			
Best Local Similarity	30.5%; Pred. No. 1.1e-19;			
Matches	100; Conservative 40; Mismatches 112; Indels 76; Gaps 15;			
QY	49 PISGV-----NGVTIVNGPKSGFGGLRSCGKCMICPCCAKGVARADEISGVAAQL 102			
DB	3 PASGVYAAQTAAGSVVL-----GLMRGGRITMLGCVCAATIRHRAEITAAVEM 54			
QY	103 GTGSVA-MVTMTHTGAGQRLHDIWTGLSAAMKAAATNGRR-----WTE----- 145			
DB	55 KRGGTAIVLVFTAHGHTDLADLMDALQGRKTPDSFRPRGAYQRLITGGTAAGRAND 114			
QY	146 -----REMYGCGGVRAVEITHGK-NGMTHVHALMFSGDV-----SENILESF 189			
DB	115 GHRADDEGIDRIGYVGMATATVATVGOINMGPHALVAVGRTGERSAQAQIVATP 174			
QY	190 --SDAMFDRTWSKIVSLGFAAPLR-----NSGGIDVAKTIGGEAD-QVLAAYLVTKIA 237			

```

Db      175 EPTGALDDEMOGHMSV-WTAAIRKRNPAFTPDNRGVDPFKLETERDANDLAEYIAKTQ 233
      238 SG--VQMEVSGDGSGRHRGNAPWEI---AVDAVG---DPOA-----LELMREFE 281
Db      234 DGRAPALTELARADLTKATGAGNAPFELLRIGDLTGCMTEDDAAGVGLJEWMLSRWHEZE 293
Qy      282 FQSGMRRAIAWSRGLRPARAGLGAELTDA 309
Db      294 RAIRGRRAIEMTRYLRQMLGLDGDTEA 321

RESULT 2
NEOR_STRCY
ID      NEOR_STRCY      STANDARD;      PRT;      437 AA.
AC      P14501;
DT      01-JAN-1990 (Rel. 13, Created)
DT      01-JAN-1990 (Rel. 13, Last sequence update)
DT      01-NOV-1990 (Rel. 16, Last annotation update)
DE      Neomycin resistance protein.
OS      Streptomyces cyanogenus.
OC      Plasmid PSB24.2.
OC      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OX      Streptomycinae; Streptomycetaceae; Streptomyces.
RN      NCBT_Taxid=80860;
RP      [1]
RA      SEQUENCE FROM N.A.
RA      Bolotin A.P., Sorokin A.V., Aleksandrov N.N., Danilenko V.N.,
RT      Kozlov Y.I.;
RL      "Nucleotide sequence of DNA of the actinomycete plasmid pSB24.2.";
      Dokl. Biochem. 283:260-263(1985).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X03756; CAA27389.1; -.
DR      EMBL; M32513; AAA8338.1; -.
DR      InterPro; IPR000989; Rep.
DR      Pfam; PF01446; Rep; 1.
KW      Antibiotic resistance; Plasmid.
FT      CONFLICT    111      111      G -> R (IN REF. 2).
FT      CONFLICT    220      220      A -> P (IN REF. 2).
FT      CONFLICT    356      356      K -> N (IN REF. 2).
FT      CONFLICT    412      412      R -> G (IN REF. 2).
SQ      SEQUENCE    437 AA; 47409 MW; 2AF29BAD04EB01B CRC64;

Query Match
Best Local Similarity 11.3%; Score 227; DB 1; Length 437;
Matches 94; Conservative 35; Mismatches 126; Indels 122; Gaps 17;

Qy      55 GTTVNGPKSGFG-GLRSCGKMGWPCGAGKVAHRADEISQVVAHOLGTGSAV-MWTM 112
Db      6 GYIFARTAGAVALGLTKCGRIWLCFVCSGDDPARPSEITEAVVSNLQCGGMAYLVTF 65
Qy      113 TMRHTAAGRLHDLMTGL-----SAAWKATNGRRKRREREMT 149
Db      66 TARHTAADRLSDLMALQGTADAETGTRRRPGAYORLITGAAM--AGDKRRKSNQEGIR 123
Qy      150 GGDGY---VRAVEITGKN-GWVHVHALIMFSGV-----SENILSPS---DMF--- 194
Db      124 GRIGITGMIRATEVIVGEGAMHPIHAILVVGRTGGRGDRIRITGTFPSBDALTEWE 183

```

```

Qy      195 DRWTSKLVSLGPAPLRNSGL-----DYRKIGERADQVLAAYLTIASGVG----- 241
Db      184 DRWRS-----GPATLARSTPGFRPTGARSQATGARATASSTSSSVRSGRORPGRVHR 239
Qy      242 -----MEVSGDGSGRHRGN-----RAWMEIAYDAVGGDPQ-----ALELM 277
Db      240 QDAGQEGEPGAGTRRPQGRPPGQHDTKTPSRIG-DLMGVPESEAAHGSLAWGLDRW 298
Qy      278 REFPGSNG-----RRAI-----AMSRGLRAR 299
Db      299 AEYETAAGAGHRVDPAPAPGPRRRRRRRGRHGRFPVDRRRRRFRGQVQWDAWMLG 358
Qy      300 AGLGAELTDAQIVQOE 316
Db      359 VGRSLDLAVVEAVEGRE 375

RESULT 3
REP_LACHI
ID      REP_LACHI      STANDARD;      PRT;      314 AA.
AC      P35857;
DT      01-JUN-1994 (Rel. 29, Created)
DT      01-JUN-1994 (Rel. 29, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Rep protein (Replication protein).
GN      REP.
OS      Lactobacillus hilgardii.
OC      Plasmid pLAB100.
OC      Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OX      Lactobacillus.
OX      NCBT_Taxid=1588;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=90264300; PubMed=2188951;
RA      Jossion K., Soetaert P., Michiels F., Joos H., Mahillon J.;
RT      "Lactobacillus hilgardii plasmid pLAB100 consists of two functional
      cassettes commonly found in other Gram-positive organisms.";
RL      J. Bacteriol. 172:3089-3099(1990).
CC      -1- FUNCTION: PRODUCES A SINGLE-STRAND NICK IN A SPECIFIC SITE OF THE
      PLASMID, AND THIS NICK RESULTS IN SINGLE-STRAND REPLICATION BY
      ROLLING CIRCLE MECHANISM.
CC      -1- SIMILARITY: TO SEVERAL PROTEINS INVOLVED IN GRAM-POSITIVE BACTERIA
      PLASMIDS REPLICATION.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M55222; AAA98163.1; -.
DR      EMBL; A14660; CAA01184.1; -.
DR      PIR; B35390; B35390.
DR      InterPro; IPR000989; Rep.
DR      Pfam; PF01446; Rep; 1.
KW      DNA replication; Plasmid.
FT      BINDING    229      DNA (BY SIMILARITY).
SQ      SEQUENCE    314 AA; 37044 MW; 951446121859D8EB CRC64;

Query Match
Best Local Similarity 6.3%; Score 127; DB 1; Length 314;
Matches 55; Conservative 50; Mismatches 107; Indels 38; Gaps 10;

Qy      75 KCMIPCCAGKVGARARAEISQV--AH-QLTGSAVAVTMTHTTAQRLHDLMTGISA 131
Db      69 KSRLLPCLNWRSMQSNQMLQVDEAHKQRTGFFLTLTAENASSENILKQEVKMGK 128
Qy      132 AKKATNGRRMTEEMWGCGGYVAIVITGKNG-WVHVHALIMFGDVSENILSPS 190
Db      129 A-----ISKLPYKPKAPKMLGIVASTETITNKGSTHQHWAHLVLFVKPTFFKDSANYTN 183

```

RESULT 5		
REP BACAM		
ID	REP BACAM	STANDARD;
AC	P13963;	PRT;
DT	01-JAN-1990	(Rel. 13, Created)
DT	01-JAN-1990	(Rel. 13, Last sequence update)

SEMPERUS PROFITUR A., CCM 1904,  
 NC STRAIN=ATCC 8014 /  
 RX MEDLINE=90022315; PubMed=25117345;  
 RA Boula A., Birtinger F., Frey L., Kammerer B., Belarbi A.,  
 RA Guyonvarch A., Hubert J.-C.,  
 RT "Structural organization of pP1, a cryptic plasmid from  
 RT *Lactobacillus plantarum* CCM 1904.",  
 TL Plasmid 22:185-192(1989)

```
CC -!- FUNCTION: PRODUCES A SINGLE-STRAND NICK IN A SPECIFIC SITE OF THE
CC PLASMID, AND THIS NICK RESULTS IN SINGLE-STRAND REPLICATION BY
CC ROLLING CIRCLE MECHANISM.
CC -!- SIMILARITY: TO SEVERAL PROTEINS INVOLVED IN GRAM-POSITIVE BACTERIA
CC PLASMIDS REPLICATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31223; AAA8164.1; -.
CC InterPro; IPR000989; Rep.
CC Pfam; PF01446; Rep; 1.
CC DNA replication; Plasmid.
CC BINDING 229
CC SEQUENCE 317 AA; 37009 MW; AD9A9B594FF73DC CRC64;

Query Match
Best Local Similarity 4.9%; Score 99; DB 1; Length 317;
Matches 58; Conservative 50; Mismatches 115; Indels 72; Gaps 12;

QY 70 LRSCG-----KGM-----ICPCAGKVGARADEISQVVAHQLG--TGS 106
DB 42 VRCGEVLFRKRIQHTKLQYQTFCHKRLCPICNMRSMKNSQLKQIIAEVAREPKR 101
QY 107 VAMVTMTMR--HTAQRGLHDLMTGLSAWKAKATNGRRM--TEREYGGCDGVRAVEIH 162
DB 102 FLFLTLTVKNAHSA---BELKVSILPALTKAFNKLTRYKRYTKNL---GYLRSTETIV 153
QY 163 GK--NGMHVHALIMFSGDVSENILSFSDAMFDRWTSKLVSLGFAPLRNSGLDVRK 220
DB 154 NEQDGSYNQHLVLFVKSYFRKSNNTYLAQAEKAKLQKALKVDPEPVHQAAYKANK 213
QY 221 IGGEADQVLAAYLVTKIASGVMEVSGSGKSRHGRAPAEIADAVAGDPOALEMRETF 280
DB 214 KETDLSIQASAEETAK-----YEKVSADYMT-----ADDERNLVVYIKNL 251
QY 281 EFGSMGRALIMSRGIRARAGLCAELTDQ-----IVGESEAPVWVA 323
DB 252 EYALAGTRQIISYG-GILKQIKODLKEDVENDGLVHVEDYTKQEMAEAEVVA 305

RESULT 7
PURL_PYRAE STANDARD; PRT; 697 AA.
AC Q82Z07,
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
DB synthase II).
GN PURL OR PA50225.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=2164397; PubMed=11792869;
RA Pitt-Gibson S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.,
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum."
RT Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
RL -!- CATALYTIC ACTIVITY: ATP + 5'-phosphoribosylformylglycinamide + L-
CC glutamine + H(2)O = ADP + phosphate + 5'-
CC phosphoribosylformylglycinamide + L-glutamate.
CC -!- PATHWAY: De novo purine biosynthesis; fourth step.
```

```
CC -!- SUBUNIT: Heterodimer of two subunits, purQ and purl.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- SIMILARITY: BELONGS TO THE FGMS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB09757; AAL2642.1; -.
CC HAMAP; MF_00420; -; 1.
CC InterPro; IPR00728; AIRS_related.
CC DR Pfam; PF00586; AIRS; 2.
CC DR Pfam; PF02769; AIRS_C; 2.
CC Purine biosynthesis; ligase; ATP-binding; Complete proteome.
CC NP_BIND 91
CC SEQUENCE 697 AA; 74373 MW; 94219332A19B8558 CRC64;

Query Match
Best Local Similarity 4.7%; Score 95; DB 1; Length 697;
Matches 52; Conservative 34; Mismatches 94; Indels 70; Gaps 10;

QY 47 GRPISGVNGYTVIVNGPKSGF-----GGLRSCGKGIICPCAGKVGARADEISQVVA 99
DB 465 GRFLAAYDSINVGSPSPSAWQFEVAEGRBAEALGVIVGKVSIVNEYKGPAPV 524
QY 100 HOLGTGSVAANTMTMRHTAQRGLHDLMTGLSAWKAKATNGRRMTEREYGGCDGVRAVE 159
DB 525 -----TVAVVVLGV-----VEDVSKNRAVMEGSGVYVGTQKEVGSSEYLH--- 568
QY 160 ITHGKNHVAHVHALIMFSGDVSENILSFSDAMFDRWTSKLVSLGFAPLRNSGLDVR 219
DB 569 -----RVHGLVAQCPSPSIDYSVEKEIAAAYVKRKAARL-----TAKDV- 606
QY 220 KIGGEADQVLAAYLVTKI--ASGVGMEV-----GSGDGKSGRHNAPW 260
DB 607 GLGG-----LAAALAKMAAAGIGADIDICAPSTTARLDYLLFESNGRFLAAGEEG- 660
QY 261 EADAVAGGD 270
DB 661 GTRVGAAGD 670

RESULT 8
ARGD_BIFLO STANDARD; PRT; 431 AA.
AC P59315,
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Acetylornithine aminotransferase (EC 2.6.1.11) (ACOAT).
DB ARGD OR BL1061.
GN Bifidobacterium longum.
OS Bifidobacteria; Actinobacteriae; Bifidobacteriales;
OC Bacteria; Actinobacteria; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karimantzou M., Snel B., Vilianova D., Berger B.,
RA Pessi G., Zwielen M.-C., Desiere F., Bork P., Delley M.,
RA Priemore R.D., Arigoni F.,
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RT Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
RL -!- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + 2-oxoglutarate = N-
CC acetyl-L-glutamate + 5-semialdehyde + L-glutamate.
CC -!- CORRECTOR: Pyridoxal phosphate (By similarity).
CC -!- PATHWAY: Arginine biosynthesis; fourth step.
```



```

CC -1- MISCELLANEOUS: May also have succinylidiaminopinelate
CC aminotransferase activity, thus carrying out the fourth step in
CC lysine biosynthesis.
CC -1- SIMILARITY: Belongs to class-III of pyridoxal-phosphate-dependent
CC aminotransferases.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; A6014729; AAN24869.1; -
CC DR HAMAP; MF 01107; -; 1.
CC DR Pfam; PF00202; aminotran_3; 1.
CC DR TIGRfam; TIGR00707; argd; 1.
CC DR PROSITE; PS00600; AA_TRANSFEX_CLASS_3; 1.
CC KW Arginine biosynthesis; Transferase; Aminotransferase;
CC PYRDOXAL phosphate; Complete proteome.
CC FT BINDING 284 284 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC SQ SEQUENCE 431 AA; 45615 MW; 1A9245B103521BCE CRC64;

Query Match 4.7%; Score 94.5; DB 1; Length 431;
Best Local Similarity 22.9%; Pred. No. 2.1;
Matches 50; Conservative 27; Mismatches 72; Indels 69; Gaps 9;

QY 134 KAATNGRRMRETEREMTGCDG--VRAVEITHGKNQMHVHVALMFGSDVSENILESPS 190
DB 8 KLTGTEBSKMLGEYSQVHNAVFGTFLAVMDHGGCAHINVDGNEYDLFLAGIAVN----- 61
QY 191 DAMFDRTSKLVSLGPAAPLRNSGLDVRKIGEDAV-----IAAYLYTKI 236
DB 62 -----SLGVAFPKM-----VKAVDAQAKVAHISNYFASEPQIELASLVLT 103
QY 237 ASGVMEVGG--DGSGRGHGNAPMEI-----AVDAVGDP-QALTELMREFEFG 283
DB 104 A---GAPEGSKVYFGNSGARGNAALAKLYGRITLPGALPSTIGKPARILMTHTGFHGR 160
QY 284 SMGRRAIAMSRLR-----ARAGIAELTDA 309
DB 161 TWGALSAATKPKGIRKPYDPLVPNIFFVRAGDKVALHDA 198

RESULT 9
CH60 SYNP7
ID CH60 SYN7 STANDARD; PRT; 544 AA.
AC P22879;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL OR WOPR OR CPN60.
OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90368561; PubMed=1975581;
RA Webb R., Reddy K.J., Sherman L.A.;
RT "Regulation and sequence of the Synecococcus sp. strain PCC 7942
RT groEL operon, encoding a cyanobacterial chaperonin."
RL J. Bacteriol. 172:5079-5088(1990).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- STRUCTURAL LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M58751; AAA27314.1; -
CC DR PIR; B36721; BVYCGT.
CC DR HSSP; P06139; IGR1.
CC DR HAMAP; MF 00600; -; 1.
CC DR InterPro; IPR001844; Chaperin Cpn60.
CC DR InterPro; IPR002423; Cpn60/TCP-1.
CC DR Pfam; PF00118; Cpn60_TCP1; 1.
CC DR PRINTS; PR00298; CHAPERONIN60.
CC DR PRINTS; PR00304; TCOMPLEXTCP1.
CC DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.
CC KW Chaperone; ATP-binding.
CC SQ SEQUENCE 544 AA; 58096 MW; 29665F9A57D190FB CRC64;

Query Match 4.7%; Score 94; DB 1; Length 544;
Best Local Similarity 21.5%; Pred. No. 2.9;
Matches 77; Conservative 37; Mismatches 98; Indels 146; Gaps 15;

QY 13 DRPVLYSSDKRGTRHRLRKLQITTSFTNACGRP-----ISGVNG 55
DB 215 DEPIRLITDKKTIGVQDLVPLVLRVANA-----GREPLVIAIDIEKALATLVVNLNG 268
QY 56 VTIVNGPKSGFGGLRSCGKGMICPCGAGVGAHRADEISQVAHOGT-----GSAVM 109
DB 269 VLVNAAVKAQFGGRRKRAAMEDIAVLVTGGQ-----TEDAKRKLDITKLDOLEKARR 321
QY 110 VTMTGHTAGQLDLMTGLSAAMKAATNGRRMRETEREMTGCDGVRAVEITHGKNQMHV 169
DB 322 ITITKQNT-----TIVAEQNEAIVKAR-----VQIRQISET----- 354
QY 170 HVHALLMFGSDVSENILESPDAMFDRTSKLVSLGPAAPLRNSGLDVRKIGEDAVLT 229
DB 355 -----ESSYDK--EKLQRLAKL-----SGVAVVKVGAATETEM 387
QY 230 -----AAYLTIASGVMEVGGSDGSGHGNAPMEIADVAVGDPQALTELMREFEFG 280
DB 388 KDRRLRLIEDALNAKKAIVEGIYVGGGTTLA-----HLAF-----QLEEM--- 427
QY 281 EFGSGRRRAIAMSRLRRAAGL-GAELTDAQIVQ-----ESAPVVAIIPAR 328
DB 428 -----ATANLGEHETGAQIVAFALTARLKRIAEVAGNCAVTSER 468

RESULT 10
MEN_DROME
ID MEN_DROME STANDARD; PRT; 796 AA.
AC P52302;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Lethal(3)malignant blood neoplasm-1 protein (lbn(83)).
GN L(3)MEN OR L(3)MEN-1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Oregon-R;
RA MEDLINE=94229362; PubMed=8174791;
RA Konrad L., Becker G., Schmidt A., Kioeckner T.,
RA Kafer-Stilliger G., Dreschers S., Bastrom J.-E., Gateff E.;
RT "Cloning, structure, cellular localization, and possible function of
RT the tumor suppressor gene lethal(3)malignant blood neoplasm-1 of
RT Drosophila melanogaster."
RL Dev. Biol. 163:98-111(1994).

```

```

CC -1- FUNCTION: REQUIRED FOR DIFFERENTIATION OF THE PHAGOCYtic BLOOD-
CC CELL TYPE, THE PLASMA CYTOTOXIC.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; MAINLY AROUND THE NUCLEUS.
CC -1- TISSUE SPECIFICITY: BLOOD CELLS AND OTHER TISSUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z47722; CAA87654.1; -.
DR PIR: S57844; S57844.
DR FlyBase: FBgn0002440; 1(3)mbn.
DR InterPro: IPR000618; Insect cuticle.
DR Pfam: PF00379; Chitin_bind_4; 1.
DR Repeat.
FT DOMAIN 225 314 2 X 37 AA REPEATS.
FT REPEAT 225 261 1-1.
FT REPEAT 278 314 1-2.
FT DOMAIN 471 512 2 X 21 AA REPEATS.
FT REPEAT 471 491 2-1.
FT REPEAT 492 512 2-2.
FT DOMAIN 576 711 7 X APPROXIMATE TANDEM REPEATS,
FT REPEAT 576 598 GLY-SER-RICH.
FT REPEAT 599 618 3-1.
FT REPEAT 619 640 3-2.
FT REPEAT 641 658 3-3.
FT REPEAT 676 694 3-4.
FT REPEAT 695 711 3-5.
FT REPEAT 711 711 3-6.
SQ SEQUENCE 796 AA; 83348 MW; 6B1D593961FEF730 CRC64;

Query Match 4.6%; Score 93; DB 1; Length 796;
Best Local Similarity 23.9%; Pred. No. 5.5;
Matches 80; Conservative 39; Mismatches 108; Indels 108; Gaps 18;

QY 5 SAHLTSG-KDRPVLVSSDPRGIRHELRPLQOITTESEFNACGRISGVNGVTIVNGPK 63
DB 522 SAKTLANTKTRPP-----HTFNHQPTLSSAT-----APGISVT-ANTPT 562
QY 64 G---SGFGIRSGCGKMTCPCCAGK--VGNRADEISQVVAHQLOTGSVAVMTWTRHT 117
DB 563 GNPSPNGGQI-AGKAPGNPQAGSGGIGTA-CAPGRKVSAGIGSGSAL----- 611
QY 118 AGQRHDLMTGLSAAMKATNGRMRTEREMVCGDGVVAVEITKCKNGHYHVALMF 177
DB 612 -----GGVSGSGKASGNGAIGSGSNAIGGATGSKASGFGFSN-----I 651
QY 178 SGDVSENIILSP-SDAMFDEMTSKVLVIGFAPLRNSGGLDVKTIIGGADQVLAAYLTKE 236
DB 652 GGVVSGSKSPGSPGSEKTIIGGADGSKASGSGSKIGGIGTGR----- 695
QY 237 ASGVNMEVSGSGDKS-----GRHGRARMEITAVDAVGGDDPPA 273
DB 696 ASGIGRIGSGRSGASATDLYKFIYILDYNGHEBTGRNGDK--QSYPAIGED--A 750
QY 274 LELMRERF---EPGSGMRRAIWSRGLRARAGLGA 305
DB 751 VQRTIETIANEFQFQPH--VSW-RKLDKALPEE 782

```

RESULT 11  
PCK CHLPN STANDARD; PRT; 600 AA.  
AC 092755;

DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Phosphoenolpyruvate carboxykinase [GTP] (EC 4.1.1.32) (PEP

```

DE carboxykinase) (phosphoenolpyruvate carboxylase) (PEPCK).
GN PCKG OR CPN0851 OR CP1018.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
CC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
CC NCBI TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=9920606; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger U., Gilmwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Uetzerback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kunata S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- CATALYTIC ACTIVITY: GTP + oxalacetate = GDP + phosphoenolpyruvate
CC + CO(2).
CC -1- PATHWAY: Rate-limiting gluconeogenic enzyme.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOENOLPYRUVATE CARBOXYKINASE [GTP]
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE01666; AAD18989.1; -.
DR EMBL: AE002559; AAF3723.1; -.
DR EMBL: AP002548; BAA99059.1; -.
DR PIR: A86597; A86597.
DR PIR: E72027; E72027.
DR PHCI-2DPAGE; Q92755; -.
DR TIGR: CP1018; -.
DR HAMAP, MF_00452; -.
DR InterPro: IPR000364; PEP_carboxykin.
DR Pfam: PF00821; PEPCK; 1.
DR ProDom: PD004738; PEPCK_N; 1.
DR PROSITE: PS00505; PEPCK_GTP; 1.
KM Gluconeogenesis; lysase; Decarboxylase; GTP-binding; Complete proteome.
FT NP BIND 216 223 GTP (POTENTIAL).
FT ACT SITE 267 267 BY SIMILARITY.
SQ SEQUENCE 600 AA; 66991 MW; 79590DCLA0A9932A CRC64;

Query Match 4.6%; Score 92.5; DB 1; Length 600;
Best Local Similarity 20.0%; Pred. No. 4.4;
Matches 82; Conservative 39; Mismatches 147; Indels 141; Gaps 18;

QY 34 LQOITTSFTFNAAC---GRPIG-GVNGVTIVNGPK-----SGFGGLRS 72
DB 160 LRSIGTKFKLCHSVGKPLSPGADVSPNCPKNSRIYHFDSDSVMSFGSGYGNAL 219

```

```

QY 73 CGKMTIPCAGKVGARHRADEISQVVAHOLGTG-----SVAMVTMT 113
DB 220 LGR-----KCVALLRLASYMAKSGOMLAHEHLIIIGITNPEKSKYKFSAPSACGKTNLML 275
QY 114 MRLTAQRLHDLMTGLSAAM-KAATNGRRTREKREYGCDDYVRAVEITHGNK-----167
DB 276 MRLPGLKXIECI--GQDIAMIRGRDRGYAVNPE-YGFFGVAFGTSERTNNALATCRS 332
QY 168 -HVHVHALLMFSGDV-SENILBSFSDAMPDRMTSKVLISG-----FAAPLRNSGG 215
DB 333 NSLFTVVALTADGDVMEGLTEQPEPLTD-WLGKRWKSGSPAAHPNSRFLAPLRQPS 391
QY 216 LDVRLKIGEGADQVLAALYLRKIASGVMEVSGDGSGRHNPAPWEIADAV--GGDPQA 273
DB 392 LDPEW-----NSPQGVPLDAIIFGG-----411
QY 274 LEHMEFEFGSMKRAIAMSRLGIRANAGIGALITDAQIVYQESAPVAVAIIPASMMMI 333
DB 412 ---RSEETPLVYELSWEHGVLTIGAGMSSTTAAIVGQLKLRHDPFAMLPFCGYNM- 466
QY 334 RTCAPYVFGELILVLE-----AGATW-----ENTR 358
DB 467 ---AYYFQHWLSFANRSLKLPKIFGVNMFAPKNNQGEFLWPGFSENLR 511

RESULT 12
AFSK_STRCO STANDARD; PRT; 799 AA.
AC P54741; OSF365; 09L002;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
GN AFSK OR SC04423 OR SC6F11.21 OR SC06.01.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=94341568; PubMed=8063104;
RA Matsumoto A., Hong S.K., Ishizuka H., Horinouchi S., Beppu T.;
RT "Phosphorylation of the afsK protein involved in secondary metabolism
in Streptomyces species by a eukaryotic-type protein kinase.";
RL Gene 146:47-56(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=96186909; PubMed=8635757;
RA Ueda K., Uneyama T., Beppu T., Horinouchi S.;
RT "The aerial mycelium-defective phenotype of Streptomyces griseus
resulting from A-factor deficiency is suppressed by a Ser/Thr kinase
of S. coelicolor A3(2).";
RL Gene 169:91-95(1996).
RN [3]
RP REVISIONS TO 239-240.
RA Matsumoto A., Hong S., Ishizuka H., Horinouchi S., Beppu T.,
RA Uneyama T.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=2196410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thompson N.R., James K.D., Harris D.B., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch B., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

```

```

RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -1- FUNCTION: INVOLVED IN THE REGULATION OF SECONDARY METABOLISM BY
PHOSPHORYLATING, ON BOTH SER AND THR, THE AFSK GLOBAL REGULATORY
PROTEIN INVOLVED IN THE CONTROL OF SECONDARY METABOLISM.
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -1- PFM: AUTOPHOSPHORYLATED AT THR AND SER RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; D45382; BA08229.2; -.
DR EMBL; AL39120; CAD5483.1; -.
DR PhosSite; P54741; -.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr_kinase.
DR Pfam; PF00069; Kinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00564; POK; 9.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR PROSITE; PS00111; PROTEIN KINASE DM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding;
KW phosphorylation; Complete proteome.
FT DOMAIN 16 271 PROTEIN KINASE.
FT NP_BIND 22 30 ATP (BY SIMILARITY).
FT BINDING 44 44 ATP (BY SIMILARITY).
FT ACT_SITE 138 138 BY SIMILARITY.
SQ SEQUENCE 799 AA; 83787 MW; 4BB9BDA169F6F5B CRC64;

Query Match 4.6%; Score 92; DB 1; Length 799;
Best Local Similarity 24.0%; Pred. No. 67;
Matches 62; Conservative 21; Mismatches 69; Indels 106; Gaps 15;

QY 118 AGORLHDLMTGLSAAM-----AATNRNR-----TERMYCQDV---154
DB 556 AGAALHD---GTAYWQDARLRALDARTCDERMSYPIGDAASCGGVPVRLQADPGVTV 612
QY 155 ---VRAVEITHGKMGVHVHALLM-----FSGDVSSENILBSFSDAMPEDRWT 198
DB 613 AAGTRVLALVAGSVHWHEFADAVFLPPTVPGPAVVGGV---YLADY-----660
QY 199 SKLVSLGPAAPLRNSGSLDVYRKIGER---DQYLAAYLRKIASGVMEVSGDGSGRH 254
DB 661 ---LGTVVALATDQDRMRATEARSSSTDPL---VAAG-HVHVGSGKGL-----704
QY 255 GNPAPWEIADAVAGDQVLAALYLRKIASGVMEVSGDGSGRH-----294
DB 705 ---YTLDAVYGTGK---NR-PQAGSDIVGAPVAGRGRIHFGSSDHLITLKADDDG 752
QY 295 GLRPAAGIGALITDAQIV 312
DB 753 RLRFWKLATGSEITGSPVY 770

RESULT 13
PERX_BRARA STANDARD; PRT; 296 AA.
AC P00434;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Peroxidase p7 (EC 1.11.1.7) (TP7).
OS Brassica rapa (turnip).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=51350;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=cv, Blanc dur d'hiver;  
 RX MEDLINE=61003872; PubMed=7408864;  
 RA Mazza G., Melinder K.G.;  
 RT "Covalent structure of turnip peroxidase 7. Cyanogen bromide  
 fragments, complete structure and comparison to horseradish  
 peroxidase C.";  
 RL Eur. J. Biochem. 108:481-489(1980).  
 RN [2]  
 RP SEQUENCE OF 32-65 AND 161-175.  
 RX MEDLINE=77138218; PubMed=849740;  
 RA Melinder K.G., Mazza G.;  
 RT "Amino-acid sequences of heme-linked, histidine-containing peptides  
 of five peroxidases from horseradish and turnip.";  
 RL Eur. J. Biochem. 73:353-358(1977).  
 CC -1- FUNCTION: Removal of H(2)O(2), oxidation of toxic reductants,  
 biosynthesis and degradation of lignin, suberization, auxin  
 catabolism, response to environmental stresses such as wounding,  
 pathogen attack and oxidative stress. These functions might be  
 dependent on each isozyme/isotom in each plant tissue.  
 CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.  
 CC -1- COFACTOR: Binds 1 protoheme IX, 1 iron(III) ion and 2 calcium  
 ions.  
 CC -1- MISCELLANEOUS: The protein shown, TP7, is the principal  
 isoperoxidase during winter in turnip.  
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. CLASSICAL PLANT  
 CC (CLASS III) PEROXIDASE SUBFAMILY.  
 CC PIR; A91094; OBNB7.  
 DR HSP; P22195; ISCH.  
 DR InterPro; IPR002016; Peroxidase.  
 DR Pfam; PF00141; peroxidase; 1.  
 DR PRINTS; PS00458; PEROXIDASE.  
 DR PROSITE; PS00435; PEROXIDASE\_1; 1.  
 DR PROSITE; PS00436; PEROXIDASE\_2; 1.  
 DR PROSITE; PS00873; PEROXIDASE\_4; 1.  
 KM Oxidoreductase; Glycoprotein; Peroxidase; Iron; Heme;  
 KM Multigene family; Calcium; Pyrrolidone carboxylic acid.  
 FT METAL 43 43 CALCIUM 1 (BY SIMILARITY).  
 FT METAL 46 46 CALCIUM 1 (VIA CARBOXYL OXYGEN)  
 FT METAL 48 48 CALCIUM 1 (VIA CARBOXYL OXYGEN)  
 FT METAL 48 48 CALCIUM 1 (VIA CARBOXYL OXYGEN)  
 FT METAL 50 50 CALCIUM 1 (BY SIMILARITY).  
 FT METAL 52 52 CALCIUM 1 (BY SIMILARITY).  
 FT METAL 170 170 CALCIUM 2 (BY SIMILARITY).  
 FT METAL 216 216 CALCIUM 2 (BY SIMILARITY).  
 FT METAL 219 219 CALCIUM 2 (BY SIMILARITY).  
 FT METAL 224 224 CALCIUM 2 (BY SIMILARITY).  
 FT ACT SITE 38 38 BY SIMILARITY.  
 FT ACT SITE 42 42 DISTAL HISTIDINE.  
 FT ACT SITE 70 70 HYDROGEN-BOUND (BY SIMILARITY).  
 FT ACT SITE 139 139 SUBSTRATE-BINDING (BY SIMILARITY).  
 FT METAL 169 169 IRON (HEME AXIAL LIGAND).  
 FT ACT SITE 241 241 HYDROGEN-BOUND (BY SIMILARITY).  
 FT DISULFID 11 91  
 FT DISULFID 44 49  
 FT DISULFID 97 292  
 FT DISULFID 176 201  
 FT MOD RES 1 1  
 FT CARBOHYD 185 185 PYRROLIDONE CARBOXYLIC ACID.  
 FT SEQUENCE 185 185 N-LINKED (GLCNAC. . .).  
 SQ SEQUENCE 296 AA; 31086 MW; 53C9CCE59B2A7937 CRC64;  
 Query Match 4.5%; Score 91; DB 1; Length 296;  
 Best Local Similarity 19.1%; Pred. No. 2.6; Mismatches 79; Indels 112; Gaps 15;  
 Matches 58; Conservative 55;  
 Db 16 PVLVSSDKRGIRHEI--RPKL-----QQTTSFTFNACGRPSGVN- 54  
 12 PNLSTVSKGVSAVSSGPRMGASILRLFFHDCFVNGCGSILDDTSTFTBQNAQPNR 71

QY 55 -----GTVIVNGPKGS---GEGGLRSG-----GKGMICPCGAKYGAHA 91  
 DB 72 NSARFTVINDIKSAVEKACPGVSCADLIATAARDSVQLGPPNNV-----KVGRDA 126  
 QY 92 DEISQVAVHQQ-----LGTSVAVVTVWTRHTAQO-----RLHDL 125  
 DB 127 KTSQAQANNSNIPAPMSLSQILSSFSAVGLSTRDVALSGAHTIGQSRCVNFRARVYN- 185  
 QY 126 WTGLSAAW-----KATNGRRKTRREMYGCGGYRAVEI---TGKGMHVHTA 173  
 DB 186 ETNINAFATLRGSCPRAGSG-----DANLAPLDINSATSPDMSYFKNLMA 233  
 QY 174 -----LLMPSGDVSENILLESFSD---AMFDRWTSKIVSGFAFLRNSGLDYRKIG 222  
 DB 234 QRELHSDQVTLFRNGSGTDSIVRGYSNPSFSNPFAMNITKGDISPLTGGSG-ELRKVC 292  
 QY 223 GQAD 226  
 DB 293 GKTN 296  
 RESULT 14  
 RECF MYCPA STANDARD; PRT; 385 AA.  
 AC Q9L7JL5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA replication and repair protein recf.  
 GN RECF.  
 OS Mycobacterium paratuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1770;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zhang Q., Kapur V.;  
 RT "Genomic organization of the Mycobacterium avium subsp.  
 paratuberculosis origin of replication region."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE RECF PROTEIN IS INVOLVED IN DNA METABOLISM; IT IS  
 REQUIRED FOR DNA REPLICATION AND NORMAL SOS INDUCIBILITY. RECF  
 BINDS PREFERENTIALLY TO SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS  
 TO BIND ATP (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE RECF FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF222789; AAF33693.1; --  
 DR HAMAP; MF\_00365; -; 1.  
 DR InterPro; IPR001238; Recf.  
 DR InterPro; IPR003395; SMC\_N.  
 DR Pfam; PF02463; SMC\_N; 1.  
 DR TIGRFAMs; TIGR00611; recf; 1.  
 DR PROSITE; PS00617; RECF\_1; 1.  
 DR PROSITE; PS00618; RECF\_2; 1.  
 KM DNA damage; DNA replication; DNA-binding; SOS response; DNA repair;  
 KW ATP-binding.  
 FT NP BIND 30 37 ATP (POTENTIAL).  
 SQ SEQUENCE 385 AA; 42001 MW; 914CBG20197IDPF CRC64;  
 Query Match 4.5%; Score 90; DB 1; Length 385;  
 Best Local Similarity 23.3%; Pred. No. 4.3;  
 Matches 78; Conservative 35; Mismatches 120; Indels 102; Gaps 17;

```

QY 55 GTTIVNGPKSGFGGLRSCGKGIWICPCAGKVGAAHRADEISQVVAHQLGTSVAAVMTMTM 114
DB 24 GRATYFIS--SNGFEKTYLLEALMY-----SSTLSHRYGTDAPIR-----AGADRAVYSTI 73
QY 115 RHTAGQELH-DLWTGLSAAMKAAKATNGRRMTEREMTGCQGVAAVEITGKNGMHVHVA 173
DB 74 VVNDGRCAVDLEIAGARKANKARINRSVPVSTRBVL-----GVLAAYVFA-----PRD 121
QY 174 LLMFSGDVSEN-----ILSEFSDA-----192
DB 122 LAIVRGDPSERRRITLDLATLRPAIAAARADYDKYLKORTALIKSLSGARRHSDRGALD 181
QY 193 MEDWTSKLVSLG--FAA--PLNNSGGLVRRKIGBADQVLA-----AYLTXTIASG 239
DB 182 TLVDWDSRLAEVGAQLMAARIDLVNQAPVEYK-----AYQLAPSGRAASIGYRSSLGAA 237
QY 240 VGMVSGSDGK-----SGRHGR-APEWETAYDAGDGPQALEMREPE-----FESMER 287
DB 238 ASAEVNAAGDRDYLEAALLAGLAARHAELEKMGCLVGPFRDLEMLGQVAKGFASHGE 297
QY 288 FAIAMSRLPARAGIAGAEITDAQIVEOEESAPYMW 322
DB 298 ---SWSLALSLR-----LAAYELRLADESDPVLL 323

```

## RESULT 15

```

ID FLB1 CAMJE STANDARD; PRT; 571 AA.
AC P56964; O9PMW1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin B.
OS FLAB OR CU1338C.
NC Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
CX NCBI_TaxID=197;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCYC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kerley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felctwell T., Holtroyd S.,
RA Uegels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000)

```

```

CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC
CC EMBL; AL139078; CAB73765.1; -
CC PIR; G81277; G81277.
CC InterPro; IPR001029; Flagellin_C.
CC InterPro; IPR001492; Flagellin_N.
CC Pfam; PF00700; Flagellin_C_1.
CC Pfam; PF00669; Flagellin_N_1.
CC PRINTS; PR00207; FLAGELLIN.
CC Flagella: Complete proteome.
CC INIT_MBT 0
CC SEQUENCE 571 AA; 59054 MW; ACE2D9B300F05053 CRC64;

```

```

Query Match 4.5%; Score 90; DB 1; Length 571;
Best Local Similarity 23.5%; Pred. No. 6.7;
Matches 67; Conservative 28; Mismatches 78; Indels 112; Gaps 16;

QY 20 SSDRRGIRHIEPLQOI-----TSETFNACGRPISGVNGVTVNGPKSGFGGLRSC 73
DB 223 SADKTGVATFTVTRGIAAIVRAGTISDTF-----AINGVTI-----259
QY 74 GKGMICPCAGKVGAAHRADEISQVVAHQLGTSVAAVMTMTMTAGQRLHDLWTGLSAAM 133
DB 260 -----GVAYEDGD-----GNGLVAALNSVKOT-----TGEVASI 290
QY 134 KAATNGRRMTEREMTGCQGVAAVEITGKNGMHVHVALLMFSGDVSEN-----ILE 187
DB 291 DA--NGQLLTSRE-----GRGIKI--DNIGGGAIRNA-----DMKENYGRSLSLVN 334
QY 188 SPSDAMPDRWTSKLVSLGFPA-----PLNNSG-----LDVKTIGEDQV--- 228
DB 335 DKKDILIS--GSNLSAGFGATQFISQASVLSRESKGRFPDANTADAMGFGSANKGVILGG 392
QY 229 ---LAAYLTK-----IASGVGMEVSGDPKSGRRHGNRAPEWETAYDA 266
DB 393 YGSVAIYMSBAGSGFSSGSGYSVSGKXNYSTGRAN---ATAISA 433

```

Search completed: November 8, 2003, 23:20:08  
Job time : 26 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: November 8, 2003, 23:09:28 ; Search time 66 Seconds  
(without alignments)  
1481.848 Million cell updates/sec

Title: US-10-007-527a-2

Sequence: 1 MTSVSAEHLGKDRPPVIVS.....HLHYRLPADYRPPVIVRK 379

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_XVIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	609.5	30.4	459	2	P94171
2	569	28.4	433	2	P94171
3	497	24.8	403	2	O69764
4	424.5	21.2	476	2	O98100
5	374.5	18.7	405	9	O88CH7
6	373	18.6	380	2	O9X3U7
7	360.5	18.0	622	2	O8G3D7
8	345	17.1	471	2	O8RIU6
9	342.5	17.1	461	2	O93141
10	326	16.3	401	2	O9R881
11	315.5	15.7	312	2	O8G8F1
12	293.5	14.6	386	2	O9ACT3
13	277.5	13.8	356	2	O52220
14	266.5	13.3	528	2	O57477
15	265	13.2	478	2	O8RMS3
16	256.5	12.8	505	2	O83016

17	235.5	11.7	481	2	O54245
18	230.5	11.5	437	2	O52205
19	194	9.7	396	2	O69069
20	169	8.4	240	2	O45282
21	161.5	8.1	261	2	O8G156
22	123.5	6.2	315	2	O56129
23	122.5	6.1	315	2	O9RNV8
24	122.5	6.1	315	2	O9RNV4
25	121.5	6.1	315	2	O9RNV5
26	119	5.9	319	2	O9ZNC0
27	115	5.7	336	2	O06460
28	114.5	5.7	315	2	O9X3N4
29	114	5.7	314	2	O30850
30	111.5	5.6	276	2	O9K4S9
31	111.5	5.6	320	2	O52769
32	111.5	5.6	320	2	O54265
33	111.5	5.6	320	2	O52776
34	111.5	5.6	333	2	O00452
35	111	5.5	339	2	O9F7Y4
36	110.5	5.5	339	2	O45447
37	109.5	5.5	315	2	O9X9N2
38	108.5	5.4	334	2	O9X9C2
39	107.5	5.4	321	2	O99135
40	105.5	5.3	743	13	O9PUT1
41	105	5.2	381	16	O915P2
42	104.5	5.2	340	2	O45450
43	104.5	5.2	340	2	O9X3Y9
44	104.5	5.2	988	5	O17434
45	104	5.2	313	2	O9X9P2

## ALIGNMENTS

RESULT 1  
ID P94171 PRELIMINARY; PRT; 459 AA.  
AC P94171;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DR 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE Rep.  
GN Rep.  
KM Rep.  
OS Arcanobacterium pyogenes.  
OG Plasmid PAP1.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Actinomycetaceae; Actinomycetaceae; Arcanobacterium.  
OX NCBI\_TaxID=1661;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BBR1;  
RX MEDLINE=98292760; PubMed=9620977;  
RA Billington S.J., Jost B.H., Songer J.G.;  
RT "The Arcanobacterium (Actinomycetes) pyogenes plasmid PAP1 is a member  
of the pU101/pU1 family of rolling circle replication plasmids.";  
RL J. Bacteriol. 180:3233-3236 (1998).  
DR EMBL; U83788; AAC46399.1; -  
DR InterPro; IPR000989; Rep.  
DR Pfam; PF01446; Rep; 1.  
SQ

SEQUENCE	459 AA; 51264 MW; 2A0A4B0A043BFB65 CRC64;
Query Match	30.4%; Score 609.5; DB 2; Length 459;
Best Local Similarity	36.3%; Pred. No. 1.2e-38;
Matches 148; Conservative 62; Mismatches 133; Indels 65; Gaps 15;	
QY	5 SAEHLGKDRPPVIVS-----KRGIRHELRPVLQOTTSETFNACGR-PISGNGV 56
DB	67 SATPLGNTVLTTPFVSVESKTKTKRSERYELRDGLAISTSVRKCGRVVAPLVSL 126
QY	57 TTVNGPFGSGFGGIRSGCGKMICPCGAGKVGARADEISGVVAHQGTG-SVAVMTWTR 115
DB	127 RAKSDGKAGAGYGYHFCGSVMACPVCSAKIAARKNDLQGVVDHAYVGMTVMTLTLOR 186



DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)



DE Replication protein.  
 GN REP.  
 OS Streptomyces natalensis.  
 OG Plasmid pSN1.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 CC Streptomyces; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=68242;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mendes M.V., Aparicio J.F., Martin J.F.;  
 RT "Complete nucleotide sequence and characterization of pSN1 from the  
 RT plasmid-producing Streptomyces natalensis that replicates by a  
 RT rolling circle mechanism."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ243257; CAB62261.1; -  
 DR InterPro; IPR000989; Rep.  
 DR Pfam; PF01446; Rep; 1.  
 KM Plasmid.  
 SQ SEQUENCE 476 AA; 51890 MW; DEB4CB83CCB5B95 CRC64;  
 Query Match 21.2%; Score 424.5; DB 2; Length 476;  
 Best Local Similarity 29.7%; Pred. No. 2,1e-24;  
 Matches 121; Conservative 53; Mismatches 148; Indels 85; Gaps 14;  
 QY 37 ITTSETNACGPRISGVNGVTVTVNGFKSGF--GGLRSCGKMWICPCGAKYGAHRADEIS 95  
 DB 1 MTTDKALSRGKALG--GGVTPKVPFGTAYLAGLATCGKVAHCPCGAKIRSATVELQ 59  
 QY 96 QV-----VAHQGTGSVAMVTMTMT 117  
 DB 60 AAGTAMEDIGGNGPKIRPAPAAKKDINDRTAEADMALEEDGLOAGGLTLTMTMHY 119  
 QY 118 AGQRLDLMTGLSAAMKAATN---GRMTEREMVCGDYVAVEITGK--NGMHVHA 173  
 DB 120 SRTTLAIVTQGDANKKALGONAGDWRKAKKDGIVGFVAMVETVGEANGMHPMHV 179  
 QY 174 LMFSGDVSENIIESSDAMFDRWTSKLVSLGPAAPLRNSG--GLDVKRIQGEADQVLAAY 232  
 DB 180 LVFEDPKLPPEQGDALVEVYEMASTALQDVAGYVDFREHGVRLDSGEGSGPR--ARY 237  
 QY 233 LVTKIAGS---VGMVEVGSDGSGRGNRAPEWIA---VDAVGSD--PQALELREFE 281  
 DB 238 LMYDQDKAAWTTAAETRTDTKAGDGRTPFELAKRVLTEDDADDBRAQVVRIMQYE 297  
 QY 282 FSGMGRRAIAMSRLRPAAGLGAEL--TDAQIVQGESAPVMAI1-----PA 327  
 DB 298 TAARGKRALYMSNGLRRLAALVELDTRDGEIAAEERQGEALAVIADPWHQHIAARRKG 357  
 QY 328 RSMNMTRTCAPIYFEGEILGLVEA--GATWENLMDHLHYRAPADVRRP 373  
 DB 358 RSLQLKAAEKGGQDQKRALVBSWGLVW-----GRDVLDP 392  
 RESULT 5  
 08SCH7 PRELIMINARY; PRT; 405 AA.  
 AC 08SCH7;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE 0-F9.  
 OS Propionibacterium phage phib5.  
 OC Viruses; ssDNA viruses; Inoviridae; Inovirus.  
 OX NCBI\_TaxID=189836;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=21886396; PubMed=11889111;  
 RA Chopin M.C., Rouault A., Ehrlich S.D., Gautier M.;  
 RT "Filamentous Phage Active on the Gram-Positive Bacterium  
 RT Propionibacterium freudenreichii."  
 RL J. Bacteriol. 184:2030-2033(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Chopin M.C., Rouault A., Gautier M.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF428260; AAL91702.1; -  
 SQ SEQUENCE 405 AA; 44762 MW; DE4B5BC4819D4F90 CRC64;  
 Query Match 18.7%; Score 374.5; DB 9; Length 405;  
 Best Local Similarity 29.4%; Pred. No. 1,2e-20;  
 Matches 111; Conservative 62; Mismatches 154; Indels 51; Gaps 14;  
 QY 8 HUSGDRPEVLVSSDKRGIRHRLPKLQITSETNACG--PISGVNGVTVV---NGPK 63  
 DB 14 HADGSNR-----RRRVMSARRTLWTGSLGFRNCGHVPRLSSGVAIVRRDSS 65  
 QY 64 G-----SGFGILSCGKMWICPCGAKYGAHRADEISQVVA--HQGTGSVAMVTMTMT 117  
 DB 66 GNVTDIGISGLQHCHSVAACPVCAQOIASQRMVLSVLSRHWARG--GRAMMTITLVHD 124  
 QY 118 AGQRLDLMTGLSAAMKAATNGRRMTEREMVCGD---YVAVEITGKNGMHVHA 174  
 DB 125 RTQSLKIVMDAVAKGMSKATNGRSDVLSGLGVYDGRPLPMLRFEVTHGSGMHVHA 184  
 QY 175 LMFSGDV---SENIIESPDAMFDRWTSKLVSLGPAAPLRN--SGLDVRIKGEADQVLA 230  
 DB 185 VILGEGARREADVADVQIPARWGSWMRAVVRQGLSGLSSEAHILTEPDAGDGLGILA 244  
 QY 231 AYLTKIASGVMEVSGSDGSGRGNRAPEWIAVAVAGD-----PQALEL 276  
 DB 245 AYFSKAAFEILT---GATKAAQGDNCPTPEI--LTLGLDNGEVGLRLSGKERRLRV 300  
 QY 277 WREFEFGSGRRRAIAMSRLRPAAGLGAELT--DQIVQGESAPVMAI1PARWMMIR 334  
 DB 301 WHEWEEFSSGRRMGYSKGLFDMIGLDSWTPSDDEITEDDDELFLVIGHEEWTLR 360  
 QY 335 TCAPVYFEGEILGLVEAGA 352  
 DB 361 D-----RGLADVLDAVA 373  
 RESULT 6  
 09X3U7 PRELIMINARY; PRT; 380 AA.  
 AC 09X3U7;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Replication protein.  
 GN REP.  
 OS Bifidobacterium breve.  
 OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.  
 OX NCBI\_TaxID=1685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=NCFB 2258;  
 RC MEDLINE=99271179; PubMed=10339821;  
 RA O'Riordan K., Fitzgerald G.F.;  
 RT "Molecular characterization of a 5.75-kb cryptic plasmid from  
 RT Bifidobacterium breve NCFB 2258 and determination of mode of  
 RT replication."  
 RL FEMS Microbiol. Lett. 174:285-294(1999).  
 DR EMBL; AF085719; AAD34709.1; -  
 DR InterPro; IPR000989; Rep.  
 DR Pfam; PF01446; Rep; 1.  
 KM Plasmid.  
 SQ SEQUENCE 380 AA; 42758 MW; D26F597563B604FA CRC64;  
 Query Match 18.6%; Score 373; DB 2; Length 380;  
 Best Local Similarity 27.9%; Pred. No. 1,4e-20;  
 Matches 97; Conservative 56; Mismatches 143; Indels 52; Gaps 11;  
 QY 58 IVNGPKSGRGGRGKRGKMWICPCGAKYGAHRADEISQVVAHQ----- 102





```

QY 72 SCGGKMTCPCCAGKGNHRADEISQVVAHQLTGTG--SVAAVTMTMTTACQRLHDLMTGLS 130
DB 2 NCGSVMAVPCVCAAVQRREREEVAKAVAMDAQGLQAAVTTLPFFTAADQAEELQOQK 61
QY 131 AAMRAATNGRMWTEREMYGCDGVYRAVEITTHGKNGMHVHALLMFSQDVSENILES 190
DB 62 TALKYLABEGSWTFFKATNGYQGLIRSELTHQNGMHHTHELMVVDAGTDADTKK-- 119
QY 191 DAMFDRWTSKLVISGFAPAPLRNSGGLDVRRKIGEBADQVLA-----AYLYK 235
DB 120 -TVLERWKTSCARAGL-----LDLNNV-----DQVAFAFAHADVKGWCTASDYLAK 165
QY 236 IAS-----GVGMEVSGSGKSGRGNRAPWEIADVAGDPQALMREFEFGSMGRRAIA 291
DB 166 QDDSRHMGVDAEIKASTKAGRAKGRKPFALLLPDQDGLKAGHRLDYAAAMKGROLF 225
QY 292 WSRGLRARAAGLGAELTDAQIWEDESAFVVAIIPASWMTMT--CA 337
DB 226 WSAGLKAKVGV--IEQSDBEVAEBEQDADDLGLHLEIEQMKLIRQAQCA 272

```

## RESULT 12

```

QY 9AGT3 PRELIMINARY; PRT; 396 AA.
AC Q9AGT3;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE Rep protein.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pml132f.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_Taxid=382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21172893; PubMed=11274136;
RA Barran L.R., Ritchot N., Bromfield B.S.P.;
RT "Sinorhizobium meliloti Plasmid pml132f Replicates by a Rolling-
RT Circle Mechanism."
RL J. Bacteriol. 183:2704-2708(2001).
DR EMBL; AF327371; AAG59887.1; -.
KW Plasmid.
SQ SEQUENCE 396 AA; 42948 MW; 89EA71F998579FE7 CRC64;

```

Query Match 14.6%; Score 293.5; DB 2; Length 396;  
 Best Local Similarity 29.4%; Pred. No. 2e-14;  
 Matches 77; Conservative 47; Mismatches 117; Indels 21; Gaps 8;

```

QY 64 GSGGGLRSGCGKMTCPCCAGKGNHRADEISQVVAHQLTGTG--SVAAVTMTMTTACQRLHDLMTGLS 121
DB 9 GARTSGTFPACDSGMLCEVCAAPRKQERDMAEVFDHVAKAFRDGQVMCTLVVRRHSRGG 68
QY 122 LHDLMTGLSAAMKATNGRMWTEREMYGCDGVYRAVEITTHGK--NMHVVHALLMFSQDV 180
DB 69 LADLRKYVQAASRRARAGAPWARKKHGIFGVIASAEVTFSSNGMHHTHALLRGT 128
QY 181 VSENIILSFSDAMFDRWTSKLVISGFAPAPLRNSGGLDVRRKIGEBADQVLAAYITKIASGV 240
DB 129 GAE--AODLGEMFVRRTLEAIOAAGSALL--DGDQVSYI--RSEKLAIFYISK--GV 178
QY 241 GME-----VSGDQKSGRGNRAPWEIADVAGDPQALMREFEFGSMGRRAIAMS 293
DB 179 GRTLDIAMEMAGQATKARSRNEGHHPET--LELASGDGKALMLLEYAAMKGVRSQVVS 237
QY 294 RGLRARAAGLGAELTDAQIWEDE 315
DB 238 PAIAALGIEADEDEERERGERE 259

```

## RESULT 13

```

QY 952220 PRELIMINARY; PRT; 396 AA.
ID Q52220

```

```

AC Q52220;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
DE Replication protein.
GN REPA.
OS Acinetobacter baumannii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_Taxid=470;
RN [1]
RP SEQUENCE FROM N.A.
RA Pu J.-F.;
RT "Sequence analysis of an Acinetobacter baumannii cryptic plasmid
RT pAB49."
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L77992; AAA9423.1; -.
DR InterPro; IPR000989; Rep.
DR Pfam; PF01446; Rep; 1.
KW plasmid.
SQ SEQUENCE 396 AA; 45656 MW; F086620D1ADEBA20 CRC64;

```

Query Match 13.8%; Score 277.5; DB 2; Length 396;  
 Best Local Similarity 27.5%; Pred. No. 3.3e-13;  
 Matches 74; Conservative 51; Mismatches 109; Indels 35; Gaps 8;

```

QY 67 FGLRSGCGKMTCPCCAGKGNHRADEISQVVAHQLTGTG--SVAAVTMTMTTACQRLHDL 124
DB 88 YGNVQIGSTWSPVCAKQKQTKQRRNELGKLGESWKTVNGSVYLLTTFSSHSPQSLKS 147
QY 125 LMTGLSAAMKATNGRMWTEREMYGCDGVYRAVEITTHGKNGMHVHALLMFSQDVSEN 184
DB 148 NLGKLRAMRAFETTRVQAIFPKLSVPHKIKGLVTVQNMHPHHVLL----- 198
QY 185 ILSEFSAMFDRWTSKLVISGFAPAPLRNS-----GLDVRRKIGEBADQVLAAYITKIA 237
DB 199 -LNEHHLRKRQDYSITELMTWKACIKSGUNPMSKRGHLDLR--GSYADQVYSKM----- 251
QY 238 SGVGMVSGSGKSGRGNRAPWEIADVAGDPQALMREFEFGSMGRRAIA 291
DB 252 -GLEDELSTKHVKKNGNGFTFDLNFISDENETIGKRP--KLFGFAISMKGAROLF 308
QY 292 WSRGLRARAAGLGAELTDAQIWEDESAFV 320
DB 309 WSRGLKRLKILGI--EKSDEBLAVETDYASI 336

```

## RESULT 14

```

QY 957477 PRELIMINARY; PRT; 528 AA.
ID Q57477
AC Q57477; O09454; O09409;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE REP protein.
GN REP.
OS Streptomyces phaeochromogenes.
OC Plasmid pUV1.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_Taxid=1923;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NRRL-B3559; PubMed=7582009;
RX MEDLINE=96036207; PubMed=7582009;
RA Servin-Gonzalez L., Samperi A.I., Cabello J., Galvan L., Suarez V.,
RA Castro C.;
RT "Sequence and functional analysis of the Streptomyces phaeochromogenes
RT plasmid pUV1 reveals a modular organization of Streptomyces plasmids
RT that replicate by rolling circle."
RL Microbiology 141:2499-2510(1995).
RN [2]

```

RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL-B3559;  
 RX MEDLINE=94052369; PubMed=8234485;  
 RA Servin-Gonzalez L.;  
 RT "Relationship between the replication functions of Streptomyces  
 RT plasmids pV1 and pV101."  
 RL Plasmid 30:131-140(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92192466; PubMed=1547947;  
 RA Denis F., Brzezinski R.;  
 RT "A versatile shuttle cosmid vector for use in Escherichia coli and  
 RT actinomycetes."  
 RL Gene 111:115-118(1992).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Denis F., Brzezinski R.;  
 RL Submitted (APR-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: ESSENTIAL FOR THE AUTONOMOUS REPLICATION OF THE PLASMID  
 CC PVI.  
 DR EMBL: U23762; AAA91008.1; -  
 DR EMBL: AF000429; AAB60866.1; -  
 DR InterPro; IPR000989; Rep.  
 DR Pfam; PF01446; Rep; 1.  
 KW Plasmid; Plasmid partition.  
 SQ SEQUENCE 528 AA; 58079 MW; DBA3464D140EDB2F CRC64;  
 Query Match 13.3%; Score 266.5; DB 2; Length 528;  
 Best Local Similarity 24.8%; Pred. No. 3.4e-12;  
 Matches 100; Conservative 45; Mismatches 137; Indels 121; Gaps 17;  
 QY 34 LOOITSETFNACGR---PISGVNGVITVNGPKSGSGFG---GLRSCGKGMICPCGAKV 86  
 DB 2 LNRVSGIDACGCGCGRRVLDPTGV---IYAKSSRGVVTIGVRCGRIMFCPCSSAI 56  
 QY 87 GAHRADDEI-SQVVAHQLTGSAVAMTMTMHTAGORLHD-----LWNG----- 128  
 DB 57 RRGRTBELTGMALRLAAGTILAVVLTARNTQTTDLSVLAALMGSPILDDKGAFLDR 116  
 QY 129 -----LSAAMKAATNGRRWTEREMV-----GDDGYRAVEIT 161  
 DB 117 SGKPRRAGAYQRMLTAPAFYGRPRARRTRKQGYVAPADGIRHRIYIGMVRRAEVT 176  
 QY 162 HG-KXGCMHVHVALMFGSDV-----SENILESDDMFPRWTSKL----- 201  
 DB 177 RSKKQGYHPLMLVFLGSELSGTPAKGDVVGHPFESTDGDWEDLREMMAGALKRAD 236  
 QY 202 ----VSLGPAAP---LRNSG-GLDVKRIAGEADQVLAAYLT-----I 236  
 DB 237 PKFPESTDCDTGCGCKGKGGVWVSIVRSADVALLLEYLTKNDGKRBRPDSVDQDLA 296  
 QY 237 ASGVGMEVSGSGKSGR-HGNRAPELA-----VDVAGDPQALB-----LMREF 280  
 DB 297 AGAAMETARLDSTKGRKRSKTPQILYRLMDIEVAGLDPDMAEGYGTPKQLRAMWAQY 356  
 QY 281 EFGSMGRRAIAMSRLRARAGIGA---ELTDAQIVEQESAPV 320  
 DB 357 EEALAGRALEWTRGLRRHVLDLDDGDDDETDLYVPEPEAPL 399  
 RESULT 15  
 QBRNS3 PRELIMINARY; PRT; 478 AA.  
 AC QBRNS3;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE Putative replicase.  
 GN REP.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteriales; Corynebacterium.  
 OC Corynebacteriaceae; Corynebacterium.

OK NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=227;  
 RA Lei C., Zheng Z., Ren Z.;  
 RT "Identification, sequencing and characterization of a new plasmid  
 RT pX2608 from Corynebacterium glutamicum 227."  
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF479770; AAL83305.1; -  
 KW Plasmid.  
 SQ SEQUENCE 478 AA; 52351 MW; 64AF16C27DDA297D CRC64;  
 Query Match 13.2%; Score 265; DB 2; Length 478;  
 Best Local Similarity 25.1%; Pred. No. 3.9e-12;  
 Matches 98; Conservative 49; Mismatches 143; Indels 100; Gaps 17;  
 QY 25 GIRHRLRPKLQOITSETFNACGR---PISGVNGVITVNGPKSGFGGLRSCGKGMICPC 82  
 DB 104 GLRGEVK-----VCGTVPLVNGVGLSVTTG-NGRVSGMWRGSRM-CPEC 148  
 QY 83 AGKVGARADEISQVVAHQLTG-SVAMTMTMHT-----TAGORLHD-----L 125  
 DB 149 RAKVAKAAAEVSAVCTALDRDLIVMTTLTASHYTKDLRAAGSLHEAVQVTTQV 208  
 QY 126 WTGLSAAAMKAATNGRRWTEREMVGGDDGYRAVEIT-----HGKGMHVHVALME- 177  
 DB 209 RTEGRAVALATGRRGADLRR--GRVGMITAQEVTTDDLIYASRTGIMHRIHILVLE 266  
 QY 178 --SGDVSENILESDDMFPRWTSKLVSLGFAPLRNSGLDVR-KIGEADQVLAAYLT 234  
 DB 267 FQBLTTRKQVANEYGERLFETWQOQCEGAGLVA--DKRGFTTATGTESALELAGIVA 323  
 QY 235 KIAS-----GVGMEVSGDGKSGRRGNRAPE-----IADVAGDPQALB- 275  
 DB 324 KGESPQAKAGEITTLBWAHPEGKKGGRSVSEQVLRNAYLCDTGLEDKKRYAR 383  
 QY 276 ---LMREFSGMGRRAIAMSRLRARAGIAGELTDAQIVEQES- SAPVMAIIPKRSNM 331  
 DB 384 LVAQMKLEVTGKGVHMLRSPGLRDLVGLGSELSDELIANTEDWEAABPAAV- 437  
 QY 332 MIRTCAVYFGEILGLVEAGATWENLRDL 361  
 DB 438 -----KWEDELEHV 446

Search completed: November 8, 2003, 23:19:23  
 Job time: 69 secs